Éric Loonis



Stats pour les sous-doués

Appliquer des recettes sans se prendre la tête

Stats pour les sous-doués

Stats pour les sous-doués

Appliquer des recettes sans se prendre la tête

ÉRIC LOONIS

ISBN : 9782493956095

Autres ouvrages de l'auteur



TABLE DES MATIÈRES

INTRODUCTION

Vous vous sentez sous-doué en statistiques et vous pensez ne jamais y arriver ? Vous vous dites que pour un « littéraire » comme vous, un simple étudiant en sciences humaines, les stats c'est trop. Et pourtant, dès le master et plus tard en thèse, il vous faudra bien en faire, si vous voulez valider vos recherches et soutenir vos mémoires. Si vous êtes thésard ou docteur, il vous faut encore des statistiques pour vos publications, nul n'y échappe ! Alors cet ouvrage est là pour vous aider.

Rassurez-vous, nous n'avons ni l'intention, ni la prétention, de vous apprendre les arcanes de la statistique et vous ne trouverez ici aucune formule mathématique. Le principe de cet ouvrage est de vous permettre, d'abord, de bien comprendre ce que vous faites lorsque vous réalisez une recherche et de connaître les principales étapes de cet art admirable de la recherche scientifique. Ensuite, il suffit de savoir exactement ce que vous voulez entreprendre et ce manuel vous donnera la recette. Inutile de savoir comment fonctionne la recette pour pouvoir s'en servir, il suffit de l'utiliser avec le logiciel de statistiques *IBM SPSS Statistics* (que nous appellerons pour faire simple « SPSS »). On n'a pas besoin de connaître la chimie pour faire de la cuisine !

Une fois la recette réalisée avec le logiciel de statistiques, le manuel vous dira simplement comment il faut lire les résultats des tests et comment les interpréter. Quoi de plus simple ?

L'ouvrage comporte trois parties :

La méthodologie : d'abord vous devez savoir à quel niveau de recherche votre étude se place : recherche empirique, expérimentale, quasi expérimentale, théorique, clinique, à cas unique ? Étape suivante : vous devez bien comprendre la différence entre variable indépendante (ce que vous manipulez) et variable dépendante (la réponse des sujets). Ensuite, il vous faut construire votre plan de recherche. Qu'allez-vous faire ? Comparer deux groupes, répéter des mesures, les deux à la fois ? Enfin vient la question des différents types de variables : quantitatives, qualitatives, nominales, ordinales, parasites…

Faire une recherche : ici on vous explique en détail et avec un tas de bons conseils, toutes les étapes d'une recherche, qui forment en même temps le plan de votre mémoire (ou de l'article destiné à la publication scientifique) : introduction, problématique, hypothèses, méthode, résultats, discussion, conclusion, références et publication. Ne vous lancez pas dans la recherche ou les statistiques sans avoir bien lu et compris cette partie de l'ouvrage, cela afin de vous épargner les erreurs classiques que commettent la plupart des étudiants (comme confondre VI et VD).

Faire des statistiques : l'ouvrage est accompagné d'une étude fictive qui vous est présentée au format Excel. Vous apprendrez à préparer le dossier Excel pour saisir vos données, calculer les scores automatiquement, rapidement, sans effort et sans erreur ; puis, préparer les différentes pages Excel en prévision des différents calculs qui seront soumis à SPSS. Aux étapes suivantes on vous dira comment paramétrer le logiciel SPSS et comment réaliser vos fichiers de données SPSS à partir du dossier Excel. C'est ensuite que vous entrerez dans le vif du sujet : statistiques descriptives, équivalence des groupes, effet des variables descriptives, statistiques inférentielles, lorsqu'on est certain que « toutes choses sont égales par ailleurs ». Chacune de ces étapes vous est expliquée très simplement, avec un support visuel (captures d'écran) tiré de SPSS. On vous dit même comment

interpréter et écrire les tests statistiques. Un enfant de 10 ans saurait le faire ! À la fin, un synoptique méthodologique reprend toutes les méthodes en une seule page synthétique.

Cet ouvrage de statistiques « faciles » est sans prétention. Il s'agit simplement de présenter et de permettre l'application de recettes statistiques simples, courantes, qui suffisent pour la plus grande partie des études en sciences humaines. Notre pari est que, là où il n'y aurait rien (lorsque l'étudiant, « sous-doué en stats », se sent incapable d'ouvrir même Excel), il y aura quelque chose, un peu de stats, pas trop, juste ce qu'il faut.

1. LA MÉTHODOLOGIE

1.1. Les types de recherche

Le graphique ci-dessous (Fig. 1) représente le « Diabolo de la recherche », deux cônes reliés par leur pointe. À partir d'une problématique, la démarche empirique enquête sur une collection de faits [F1] [F2] [F3]…, dont elle va tirer, par convergence, certaines lois générales.

La démarche théorique va s'emparer de ces lois générales pour en tirer un modèle (une théorie) et imaginer un plan expérimental. Ce plan représente le moyen rationnel de tester un certain nombre d'hypothèses [H1] [H2] [H3]…, par divergence à partir du modèle.

Les résultats de l'expérience seront analysés à un niveau statistique (effet de répétition qui donne la preuve du modèle et de la loi générale qu'il organise), puis discutés et des conclusions en sont tirées, qui valideront ou non les hypothèses ou l'ensemble du modèle.

Tout cela donne généralement lieu à publication pour la communauté des chercheurs (et pour la société qui finance les chercheurs) et vient enrichir la problématique. Cet enrichissement sera à son tour repris dans de nouvelles recherches et ainsi de suite…

De temps à autre, il est bon de relever un cas particulier, par exemple [F1], dans la collection de faits et de l'étudier en profondeur comme référence, dans une confrontation directe

avec le modèle, cela pouvant illustrer et renforcer le modèle, ou au contraire, le mettre en doute, ou l'enrichir par le jeu des exceptions.

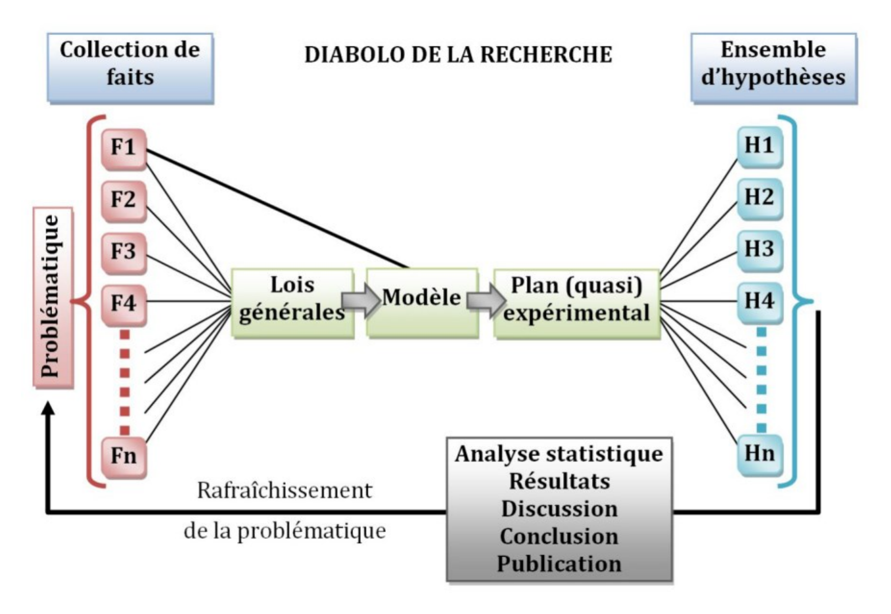


Figure 1 : Le diabolo de la recherche.

La recherche empirique

L'activité de recherche débute généralement par une démarche empirique, dite aussi convergente. Une enquête, des observations, produisent une série ou collection de faits. Ces faits sont organisés et on tente d'en tirer des récurrences, des invariants, ou des variances systématiques, à partir desquels (démarche convergente) on tirera une loi générale. Telle est la démarche empirique, qui prend uniquement appui sur ce qui est observé dans la collection du particulier pour en tirer des lois générales. En principe, la démarche empirique ne réalise aucune « manipulation » (agir sur des variables, choisir des groupes à comparer…), tout au plus le chercheur va-t-il cibler telle ou telle population afin de cerner sa problématique.

Le mémoire ou l'article de niveau empirique est introduit par une revue de la littérature qui rend compte des tentatives antérieures d'observation et d'enquête et de leurs résultats. Cette

revue vient justifier la nécessité de la présente étude. Puis, la méthode-observation d'enquête est présentée, avec éventuellement la référence à des instruments validés repris pour cette étude. Les résultats de l'enquête sont présentés sous forme de tableaux statistiques (ici il s'agira essentiellement de statistiques descriptives). Une discussion termine l'article, où l'on dégage les grandes tendances montrées par les résultats, où l'on présente la validité écologique de l'enquête, au sens où elle révèlerait quelques lois générales qui pourraient faire l'objet d'hypothèses dans le cadre de recherches plus expérimentales.

Recherche théorique

À partir des invariants, des récurrences que le chercheur a cru tirer de ses observations et enquêtes, il a pu formuler des lois générales, des principes, qu'il va organiser et systématiser suivant une construction théorique, imaginaire, un modèle. Ce modèle, qui met en œuvre des concepts nouveaux, représente une version imaginée des phénomènes réels étudiés, qui propose d'expliquer les lois observées par la supposée mise en jeu de mécanismes, de processus cachés. Ces suppositions feront l'objet d'hypothèses pour la recherche expérimentale.

Le mémoire ou l'article de niveau théorique commence par une revue de la littérature qui rend compte des différents modèles existants, en rapport avec la problématique choisie, afin de montrer leur pertinence, mais aussi leurs limites. On avancera aussi les nouveaux faits qui peuvent rendre caducs les modèles anciens et nécessiter la construction d'un nouveau modèle. On doit pour cela rappeler brièvement les faits d'observation et les lois qui en ont été tirées. De là, on forgera de nouveaux concepts qui seront articulés en un modèle explicatif. Dans la discussion, on proposera un certain nombre d'hypothèses à tester à partir de ce nouveau modèle et on en précisera les limites.

Recherche expérimentale

À partir d'un modèle, d'une théorie, un certain nombre d'hypothèses (ce qui correspond à une démarche divergente)

sont posées, qui vont demander à être testées. Pour cela, un plan d'expérience est monté. À la différence de la simple observation-enquête sur le terrain, qui se borne à prendre des mesures pour en faire une analyse, le plan expérimental cherche à manipuler des variables, à en contrôler d'autres. Classiquement la manipulation porte sur les conditions environnementales ou d'action des sujets ; mais la « manipulation » de base de toute recherche expérimentale est le découpage de la population étudiée en au moins deux sous-groupes que l'on va comparer (par exemple, le groupe d'étude et un groupe témoin ou groupe de contrôle). Ce que manipule le chercheur représente les variables indépendantes. Les hypothèses sont des prédictions que le chercheur fait quant à l'issue de son expérience, en fonction du modèle sur lequel il s'appuie (comment la variable indépendante, manipulée par le chercheur – comme constituer des groupes – va influencer la variable dépendante que représente la réponse des sujets à une expérience – qui peut être une réponse à un questionnaire rétrospectif, ou une réponse comportementale dans une situation). Les résultats de l'expérience vont valider ou invalider les hypothèses.

Recherche quasi expérimentale

Ce qui est dit ci-dessus vaut aussi pour la recherche quasi expérimentale ; et ce qui est dit au paragraphe suivant vaut aussi pour la recherche expérimentale. Ce qui distingue l'expérimental du « quasi » expérimental, c'est que dans le premier cas vous faites votre recherche dans un laboratoire (au sens physique du terme) : vos sujets viennent vous voir, ils sont placés dans une pièce spéciale, soumis à des dispositifs adaptés et vous contrôlez un maximum de variables parasites. Une recherche est quasi expérimentale (c'est le cas de la plupart des recherches en sciences humaines réalisées par les étudiants) lorsque vous allez à la rencontre de vos sujets, hors de tout laboratoire, et que vous êtes soumis à un grand nombre de variables parasites (comme les conditions d'évaluation, le degré d'authenticité ou de véracité des réponses, une évaluation rétrospective, un défaut de contrôle

de certaines variables démographiques, organismiques, un défaut d'appariement des groupes, etc.).

Le mémoire ou l'article de niveau expérimental ou quasi expérimental doit commencer par bien poser le modèle choisi pour l'étude et justifier ce choix (pertinence du modèle en regard des résultats expérimentaux antérieurs). Une revue de la littérature rendant compte des expériences antérieures est alors nécessaire pour pouvoir justifier la présente expérience et sa nécessité. Puis, la méthode expérimentale est présentée, avec ses différents instruments de mesure et leurs références. Le plan d'expérience est décrit avec précision (il doit pouvoir être reproduit), les variables sont précisées : celles qui conditionnent l'expérience (variables indépendantes), celles qui sont mesurées (variables dépendantes) et les variables parasites qu'il faudra envisager (des variables un peu trop « indépendantes »). Les hypothèses sont exposées comme autant de prédictions quant aux résultats. Enfin, après une brève présentation statistique de la population étudiée, les résultats sont présentés sous forme de tableaux ou graphiques statistiques (résultats comprenant dans l'ordre : équivalence des groupes, effets des variables statistiques sur les variables dépendantes et statistiques inférentielles – validation/invalidation de vos hypothèses). La discussion reprend les résultats en regard des hypothèses posées et valide ou invalide ces hypothèses. Les variables parasites sont examinées. À partir de là, le modèle est discuté, voire remis en question. En conclusion, de nouveaux concepts peuvent émerger, un nouveau modèle être évoqué, ou de nouvelles hypothèses formulées.

Recherche clinique

La démarche clinique, l'étude de cas, s'apparentent à la démarche empirique au sens où le chercheur se contente d'observer, de glaner des mesures. La différence va porter sur l'intention finale de la démarche. Alors que la démarche empirique cherche à multiplier les observations sur un grand nombre de sujets afin de pouvoir en tirer des singularités statistiques, la démarche

clinique borne ses observations à un ou quelques cas, qui seront observés « en profondeur ». Le but n'est pas la recherche de la singularité statistique, mais l'illustration. Il s'agit de saisir quelques éléments de la réalité et d'explorer en détail ces éléments afin de soutenir un modèle par des exemples précis.

Le mémoire ou l'article de niveau clinique commence par rendre compte d'observations, d'études de cas antérieurs, glanés dans la littérature scientifique. Il faut aussi préciser dans quel modèle on s'inscrit, car le sujet sera observé différemment selon le modèle choisi. On présentera ensuite le contexte et la méthode d'observation, les instruments utilisés. Le cas est enfin présenté : son identité démographique, son histoire, son état actuel. On pourra donner les résultats des instruments validés (sur des populations de référence) utilisés (tests, échelles…). L'étude clinique peut aussi prendre un petit air « expérimental » en comparant deux ou plusieurs cas extrêmes (entre différents troubles, entre le normal et le pathologique, entre différents types de personnalité…), sinon que l'exiguïté de la population interdira toute inférence statistique. Généralement, on conclura l'article clinique par une discussion sur les limites de l'observation de ce(ces) cas, sur les limites du modèle choisi, sur sa pertinence, qui peut être renforcée ou affaiblie.

À noter que tout utile qu'elle soit pour exemplifier et illustrer l'observation, la recherche en sciences humaines ne peut se contenter de recherches cliniques. La validité d'un modèle et d'hypothèses ne peut être testée qu'à partir d'études expérimentales (ou quasi expérimentales). Il n' y a pas non plus à s'affronter sur ces méthodes, elles sont complémentaires dans la démarche globale de la recherche scientifique en sciences humaines.

1.2. Les variables indépendantes et dépendantes

Pour beaucoup d'étudiants, cette distinction entre variable indépendante (VI) et variable dépendante (VD), est souvent un gros problème qu'ils peuvent traîner jusqu'au jour de soutenance

de leur thèse. Aussi, si vous ne deviez comprendre qu'une seule chose en méthodologie, c'est bien cette différence entre VI et VD.

La première chose à savoir est que ce qui distingue une VI d'une VD n'est pas dans la nature de la variable. Si vous prenez, par exemple, le niveau d'anxiété mesuré avec une échelle d'anxiété, cette variable « niveau d'anxiété » peut être une mesure appliquée sur deux groupes « alcoolique » et « non alcoolique » et être ainsi une variable « dépendante » (VD) de l'alcoolisme (par exemple, les alcooliques sont plus anxieux que les non-alcooliques, en réponse à l'échelle d'anxiété).

Cependant, vous pourriez aussi bien mesurer l'anxiété dans une population de buveurs en général et dichotomiser cette variable en deux groupes « anxieux » et « non-anxieux », l'anxiété devenant alors une VI. À partir de là, vous faites passer une échelle de consommation d'alcool à vos sujets, ce qui va représenter la VD (et non plus la VI) : le résultat pouvant être que les individus les plus anxieux sont ceux qui boivent le plus d'alcool. Le tableau ci-dessous montre clairement que les protocoles 1 et 2 voient alcoolisme et anxiété échanger leurs rôles en matière de VI et VD :

	Protocole 1	**Protocole 2**
VI	Alcoolisme	Anxiété
VD	Anxiété	Alcoolisme
Hypothèse	L'alcoolisme influence l'anxiété ou l'anxiété dépend de l'alcoolisme.	L'anxiété influence l'alcoolisme ou l'alcoolisme dépend de l'anxiété.
Exemple	Une étude pourrait montrer qu'après des effets positifs anxiolytiques, l'alcool consommé à long terme entraîne une augmentation de l'anxiété sociale.	Une étude pourrait montrer que les personnes qui s'alcoolisent le plus sont aussi celles qui ont les niveaux d'anxiété les plus élevés.

Vous allez vous demander : mais où est la vérité ? En science, et surtout en sciences humaines, il n'y a jamais de vérité absolue ni de causalité linéaire. Certes, le fait de boire trop d'alcool peut jouer sur le niveau d'anxiété (effets cérébraux de l'alcool), mais on peut aussi dire, en sens inverse, que plus on est anxieux et plus on éprouve le besoin de se soulager en buvant. Les choses vont dans les deux sens, la causalité est circulaire, non linéaire (un seul sens).[1] De là, on peut imaginer faire de la recherche dans un sens ou dans l'autre, c'est pour cela que VI et VD sont interchangeables.

VI et VD sont donc attribuées aux mesures d'une façon arbitraire, selon le projet de recherche du chercheur.

L'erreur qui tue !

Il s'agit d'un raisonnement très commun, basé sur une confusion sémantique, qui entraîne une mauvaise perception de VI et VD. Un exemple : un étudiant étudie un trait de personnalité, la « recherche de sensations », chez des adolescents fumeurs et non-fumeurs. Concrètement, que fait l'étudiant ? Dans ce cas précis, il prend deux groupes d'adolescents : des « fumeurs » et des « non-fumeurs », qu'il distingue avec une échelle d'évaluation du tabagisme, et il leur fait passer une échelle de recherche de sensations (évaluation du trait de personnalité). Voici son hypothèse générale :

Le trait de personnalité « recherche de sensations » chez les adolescents, affecte le degré et l'intensité de leur dépendance tabagique.

À la suite de cela, notre étudiant présente ses variables et annonce que le trait de recherche de sensations est la VI et le

[1] C'est d'ailleurs pour cela que les alcooliques, et tous les addictés en général, se retrouvent enfermés dans un cercle vicieux (la spirale addictive) qui fait que leur « solution » à la souffrance psychique entraîne elle-même une souffrance psychique secondaire.

tabagisme la VD. Car le raisonnement de l'étudiant est le suivant : pour lui « plus les scores de recherche de sensations sont élevés et plus l'addiction au tabac sera sévère ». C'est cette façon de dire les choses, qu'il tire directement du verbe « affecter » présent dans son hypothèse, qui le conduit à voir les réponses à l'échelle de personnalité comme la VI et le tabagisme comme la VD.

Pourtant, plus loin, au moment où il présente l'instrument de mesure de la recherche de sensations, l'étudiant dira que : « Nous nous attendons à ce que les fumeurs aient de plus hauts scores de recherche de sensations que les non-fumeurs ». Ici nous voyons bien que ce que fait l'étudiant est de comparer fumeurs et non-fumeurs du point de vue de la réponse de trait « recherche de sensations », et non de comparer la réponse tabagique de groupes de sujets selon leur niveau de *sensation seeking*. Voici comment on peut résumer cela dans un tableau :

	Ce que l'étudiant croit faire	**Ce qu'il fait réellement**
VI	Recherche de sensations	Tabagisme
VD	Tabagisme	Recherche de sensations
Hypothèse	La recherche de sensation a une influence sur (affecte le) tabagisme.	Le tabagisme a une influence sur la réponse à l'échelle de recherche de sensations.

D'où vient la confusion ?

En fait, ce sont les verbes indiquant une causalité, une influence, comme « affecter », « influencer », « agir sur » et leurs équivalents, qui posent problème. À cause de ces verbes, deux types de causalité sont confondus : la causalité théorique et la causalité expérimentale. Nous l'avons expliqué plus haut : la

causalité théorique, celle du modèle, n'est pas linéaire, mais le plus souvent circulaire. Un trait de personnalité affecte une conduite addictive, mais en retour, une conduite addictive, par son intensité et sa durée, peut aussi affecter la personnalité. Du côté de la causalité expérimentale, le chercheur choisit l'un ou l'autre sens dans cette circularité, car il ne peut pas travailler en disant une chose et son contraire. Il ne peut pas faire non plus, en même temps, deux choses contradictoires. Le chercheur est contraint de choisir la variable qu'il va manipuler, ce qui définit les variables de réponse des sujets en fonction de cette manipulation. C'est là la base de tout protocole expérimental.

Notre étudiant a en tête la théorie (la recherche de sensations a une influence sur le tabagisme) et se mélange les pinceaux en faisant « le tabagisme a une influence sur la réponse à l'échelle de sensations ». Son erreur a été d'utiliser un verbe d'action causale dans son hypothèse générale. Au lieu de dire :

Le trait de personnalité « recherche de sensations » chez les adolescents, affecte le degré et l'intensité de leur dépendance tabagique.

Il aurait mieux valu qu'il dise :

Il existe un lien entre la recherche de sensations et le tabagisme.

En omettant de donner un sens à cette « liaison » (ne parler simplement que de « lien »), son esprit serait resté libre d'appréhender ce qu'il allait réellement faire. Or, il n'a pas fait des groupes de recherche de sensations (suivant le degré de recherche de sensations), pour en mesurer la sévérité du tabagisme, mais il a fait deux groupes, fumeurs et non-fumeurs, pour en mesurer le trait de personnalité. Ce qu'il a manipulé,

c'est la variable « tabagisme », qui est la VI, et non la variable « trait », qui est bien la réponse, la VD. Un protocole de recherche représente toujours la mesure d'une réponse selon une variable manipulée, c'est la causalité expérimentale, qu'il ne faut pas confondre avec la causalité théorique.

À titre général, VI et VD peuvent être définie ainsi (voir le tableau ci-dessous) :

Variable Indépendante	**Variable Dépendante**
La VI (variable indépendante) est la variable manipulée par l'expérimentateur.	La VD (variable dépendante) est la réponse des sujets aux instruments de mesure, réponse qui « dépend » de la variable manipulée, la VI.
La variable indépendante est encore appelée « explicative » ou « invoquée ».	La variable dépendante est encore appelée « expliquée » ou « provoquée ».
C'est ce que manipule le chercheur.	C'est la réponse comportementale des sujets, attendue et produite.
Exemples : – sélection d'une population ; – découpage en sous-groupes* ; – choix et regroupement des variables sociodémographiques et organismiques (âge, sexe…) ; – conditions expérimentales (situations, épreuves, tâches)	**Exemples :** – réponses comportementales générales, spécifiques ; – réponses aux instruments de mesure, tests, échelles, questionnaires.

* Attention : il arrive souvent que ce « découpage en sous-groupes » soit objectivé par l'utilisation d'un instrument de

mesure (par exemple, une échelle d'évaluation du tabagisme), mais cela ne veut pas dire qu'il s'agit d'une VD ! Vous utilisez simplement un instrument pour opérationnaliser votre manipulation (construction des groupes) au niveau de la VI (mais dans un autre protocole de recherche vous pourriez très bien utiliser l'échelle de tabagisme comme « réponse » expérimentale et, dans ce cas, elle correspondra bien à une VD – la VI pouvant être, par exemple, le trait de personnalité).

VI et VD sont articulées l'une avec l'autre dans les hypothèses, qui s'organisent hiérarchiquement ainsi :

- **Hypothèse générale (ou théorique) :** à partir de la problématique et du modèle qui cherche à l'expliquer, le principe, la loi générale, organisés par le modèle, représentent l'hypothèse générale ou théorique. Dans cette hypothèse, il est rare que les variables soient mentionnées. On y parle plutôt de principes, d'entités générales, qui forment un système, remplissent des fonctions, fonctionnent entre elles, sans que soit clairement posée une causalité linéaire (on se borne à parler de « liens », de « liaisons »).
- **Hypothèses de travail :** ici, les variables étudiées apparaissent et sont articulées entre elles comme VI et VD, mais sous une forme descriptive générale de l'influence de VI sur VD, définies selon des conditions et données générales.
- **Hypothèses opérationnelles :** suivant le plan d'expérience (le protocole de recherche), VI est précisé (groupe, situation, traitement…) et une prédiction est faite de l'influence de cette VI (il peut y en avoir plusieurs) sur la VD (il peut y en avoir plusieurs) qui est elle-même précisée (réponse à une tâche, une épreuve, un apprentissage, un test, un questionnaire, une performance…).
- **Hypothèses statistiques :** si les réponses (VD) présentent des valeurs normalisées (d'autres expériences ont permis d'établir des normes de réponses), il est possible de reprendre de façon chiffrée, quantifiée, les hypothèses de travail en

prédisant des valeurs attendues. Sinon, on prédit des niveaux de réponse différenciés, des écarts de réponse, entre les sous-groupes définis par la VI.

Il n'y a normalement qu'une seule hypothèse générale (théorique) pour un protocole ; de cette hypothèse générale découle le plus souvent plusieurs hypothèses de travail, qui définissent des orientations dans le protocole, et chaque hypothèse de travail donne naissance à plusieurs hypothèses opérationnelles, en général une par instrument de mesure utilisé. Les hypothèses opérationnelles, surtout si elles sont chiffrées (statistiques) et si elles doivent poser des prédictions sur plusieurs groupes, ne sont pas obligatoirement rédigées de façon littéraire, mais peuvent avantageusement être présentées sous la forme de tableaux, qui annoncent les valeurs, ou les niveaux attendus.

Les variables intermédiaires

Dans un modèle simple de type S-R, les variables indépendantes (les stimuli S) et leurs influences sur les variables dépendantes (les réponses des sujets R) sont suffisantes et ce qui se passe dans la « boîte noire » de l'organisme est ignoré. Cependant, il est possible d'adopter une position plus complexe et considérer des variables, entre S et R, rattachées à l'organisme, le schéma devenant alors S-O-R.

Ces variables de l'organisme, dites « variables intermédiaires », correspondent, par exemple en psychologie, à des processus psychologiques ou à des activités intérieures des sujets, processus et activités déclenchés par les VI et dont la présence et les caractéristiques sont révélées par inférences, à partir de l'observation des comportements manifestes. Quelques exemples : la motivation, les processus cognitifs.

Non observables directement, donc non mesurables, les variables intermédiaires doivent être définies en regard de leurs référents empiriques. Cependant, elles gardent toujours le statut de construits hypothétiques. Elles se situent au-delà de la

statistique et ne font qu'aider la discussion des résultats obtenus entre VI et VD.

Les variables parasites

Que la recherche soit confinée au sein d'un laboratoire, en étant purement expérimentale, ou qu'elle se déroule dans la « nature » étant alors quasi expérimentale, il apparaît toujours des variables qui sont non pertinentes pour les objectifs de la recherche, qui la « parasitent » et qu'il convient de contrôler. Ces variables sont au même rang que les variables indépendantes manipulées par le chercheur, au sens où elles peuvent influencer les variables dépendantes, les réponses des sujets. Plusieurs types de variables parasites sont à considérer :

- **Variables démographiques :** lieu de résidence, situation familiale, catégorie socioprofessionnelle d'appartenance, niveau d'études, revenus, type culturel, etc.
- **Variables organismiques :** sexe, âge, ethnie, apparence, qualités physiques ou physiologiques, état de santé, etc.
- **Variables historiques :** événements significatifs de la vie, conditions d'éducation, trajectoire, par exemple, d'une addiction, d'un comportement à problème, maladies passées, traitements passés, emprisonnement, placement familial, etc.
- **Variables de dispositif :** il s'agit des variables parasites qui appartiennent au dispositif de la recherche elle-même, comme le sexe ou la race de l'expérimentateur s'il entre en contact directement avec les sujets étudiés, l'environnement et l'ambiance au cours de l'expérience, l'état des sujets au cours de l'expérience (physiologique et mental), certains détails expérimentaux comme la disposition des personnes, de l'éclairage, des instruments de contrôle, la technique et la durée de passation des questionnaires et tests, etc.

Les variables parasites apparaissent systématiquement dans la question de l'équivalence des groupes, mais encore dans la

constance des conditions expérimentales à chaque mesure répétée. Comme le chercheur veut faire apparaître au mieux le seul effet des variables indépendantes qu'il a choisies, selon la problématique dont il traite, il souhaite éliminer, neutraliser ou contrôler les autres variables qui pourraient jouer sur les résultats. Plusieurs moyens sont à sa disposition, qui s'appliquent selon les types de variables à contrôler :

- **Maintien à un niveau constant :** vous faites en sorte que le niveau de la variable parasite reste constant. Par exemple, si vous ne pouvez éliminer le bruit au cours de mesures expérimentales, vous essaierez de maintenir constant le niveau de ce bruit. Si le sexe des sujets peut jouer sur leur réponse, vous pouvez ne choisir que des sujets de sexe féminin ou masculin, afin de bien mettre en évidence l'effet de votre variable indépendante.
- **Variation systématique :** il s'agit de repérer des niveaux représentatifs de la variable parasite et de les produire d'une façon volontaire, systématique, en respectant une distribution équivalente pour chaque niveau. Par exemple, vous essayez qu'il y ait le même nombre de sujets dans chaque catégorie socioprofessionnelle, pour chaque niveau intellectuel ou niveau d'études, autant d'hommes que de femmes, et ainsi de suite. Pour une variable de dispositif, comme le degré d'abstraction de mots à mémoriser, vous pouvez considérer 5 niveaux d'abstraction et répartir également les mots à mémoriser sur ces 5 niveaux.
- **Variation au hasard :** vous pouvez laisser la variable parasite jouer librement, en la considérant comme une variable naturelle. Cependant, il est nécessaire que cela s'applique sur un nombre important de sujets afin que le hasard, justement, ne détermine des niveaux trop particuliers de la variable. Il est aussi bon de contrôler la « normalité » de cette variable, à savoir si la distribution de ses niveaux est statistiquement normale, que les valeurs centrales ou modérées sont plus

fréquentes que les valeurs extrêmes. Par exemple, pour la variable CSP (catégorie socioprofessionnelle), avec un grand nombre de sujets, ceux-ci peuvent se répartir au hasard et d'une façon équilibrée entre toutes les catégories et vous pourriez vérifier la fréquence de répartition afin de contrôler s'il n'y a pas une sur- ou sous-représentation d'une catégorie sur les autres.

- **Élimination :** plus facile à dire qu'à faire, l'élimination de certaines variables parasites, surtout celles liées au dispositif expérimental, est parfois une solution. Par exemple, le sexe de l'expérimentateur, qui peut influer sur la réponse des sujets, peut faire l'objet d'une variation systématique (il faut qu'il y ait le même nombre de sujets pour les deux genres d'expérimentateurs), ou bien vous éliminez complètement l'expérimentateur en plaçant les sujets face à un ordinateur avec lequel ils passent leurs tests.

Bien entendu, dans le cadre des expériences sur l'être humain, on ne peut jamais tout contrôler . Il faut simplement faire le maximum et être attentif à éviter les erreurs grossières. Préférer aussi certaines méthodes à d'autres. Par exemple, si possible éviter les mesures rétrospectives, car alors la variable « mémoire » peut vous jouer des tours. Et si vous devez faire du rétrospectif, évitez alors des empans temporels trop longs ou trop éloignés dans le passé.

Une variable n'est véritablement parasite que selon son degré de pertinence par rapport à votre protocole de recherche. Si, par exemple, les travaux antérieurs ont largement montré que la variable « situation familiale » n'avait aucune influence sur ce que vous voulez mesurer, alors vous pouvez sans doute négliger le contrôle de cette variable (sauf si vous escomptez remettre en question les travaux antérieurs…).

Vous ne devez pas oublier que si le contrôle des variables parasites augmente la validité interne d'une recherche, il peut en diminuer la validité externe ou écologique. Par exemple, si vous

contrôlez le sexe en ne choisissant que des hommes, vous ne pourrez prétendre que vos résultats s'appliquent aussi aux femmes.

Enfin, vous verrez plus loin que les variables parasites sont contrôlées de façon statistique, par des tests destinés à mesurer le degré de leur influence, afin de s'assurer de respecter le principe expérimental fondamental du « toutes choses égales par ailleurs ».

1.3. Les plans de recherche

1.3.1. Plans de recherche à groupes indépendants

Le principe de toute recherche en sciences humaines consiste à soumettre des groupes de sujets aux différents niveaux d'une ou plusieurs variables indépendantes afin de mesurer les effets de cette(ces) variable(s) indépendante(s) sur la réponse comportementale des sujets qui y sont soumis.

Par exemple, une variable indépendante peut être une cure de sevrage alcoolique. Les deux niveaux de cette variable sont : 1) suivre une cure de sevrage et 2) ne suivre aucune cure de sevrage. On divise la population des sujets en deux groupes équivalents par ailleurs et l'on soumet le premier groupe à la condition (1) de la variable indépendante (suivre la cure), tandis que le second groupe est laissé à sa vie habituelle, le niveau (2) de la variable.[2] Ensuite, on entreprend un certain nombre de mesures, comme le taux de consommation d'alcool par jour, les niveaux de dépression et d'anxiété. Ces mesures sont la réponse des sujets, c'est-à-dire les variables dépendantes, qui vont dépendre, selon les hypothèses posées, du niveau de la variable indépendante suivre ou non une cure de sevrage.

[2] Pour des raisons éthiques, en pratique, ces sujets peuvent être mis en « liste d'attente pour une cure », bien que souvent cette mise en liste d'attente peut avoir son propre effet placebo, qui peut lui-même avoir un impact sur les réponses des sujets…

L'articulation organisée entre variable indépendante et variable dépendante s'appelle un « plan de recherche ».[3] Il existe trois grandes catégories de plans de recherche. Voyons la première : les plans de recherche à groupes indépendants.

Des groupes différents de sujets sont soumis chacun à un niveau de la variable indépendante. Le nombre de groupes va dépendre du nombre de variables indépendantes et du nombre de niveaux de chacune des variables indépendantes. Pour trouver le nombre de groupes nécessaire, il suffit de multiplier entre eux les nombres de niveaux de chaque variable indépendante (VI). Quelques exemples :

Une VI à deux niveaux

VI a1	**Gr. 1**	*Nombre de groupes = 1 variable x 2 niveaux = 2* Ex. : la variable « cure de sevrage alcoolique » (VI a) a deux niveaux : un groupe (Gr. 1) est soumis à la cure (VI a1), l'autre (Gr. 2) non (VI a2).
VI a2	**Gr. 2**	

Une VI à quatre niveaux

VI a1	**Gr. 1**	*Nombre de groupes = 1 variable x 4 niveaux = 4* Ex. : la variable « privation de sommeil » (VI a) a quatre niveaux : les quatre groupes (Gr. 1, Gr. 2, Gr. 3, Gr. 4) sont respectivement soumis à 2 heures (VI a1), 4 heures (VI a2), 6 heures (VI a3) et 8 heures (VI a4) de privation de sommeil.
VI a2	**Gr. 2**	
VI a3	**Gr. 3**	
VI a4	**Gr. 4**	

[3] On peut aussi dire « protocole ».

Deux VI à deux niveaux

	VI b1	VI b2
VI a1	Gr. 1	Gr. 2
VI a2	Gr. 3	Gr. 4

Nombre de groupes = 2 variables x 2 niveaux = 4

Ex. : la variable « privation de sommeil » (VI a) a deux niveaux : avec privation (VI a1) ou sans (VI a2) ; la seconde variable est le « sexe » (VI b), avec naturellement deux niveaux : féminin (VI b1) et masculin (VI b2). On a donc : un groupe de femmes privées de sommeil, l'autre non (Gr. 1 et Gr. 3), et la même chose pour les deux groupes d'hommes (Gr. 2 et Gr. 4), ce qui permet de contrôler, par exemple, l'effet sur la performance intellectuelle, à la fois de l'appartenance au sexe et de la privation ou non de sommeil.

Une VI à quatre niveaux et une VI à trois niveaux

	VI b1	VI b2	VI b3
VI a1	Gr. 1	Gr. 2	Gr. 3
VI a2	Gr. 4	Gr. 5	Gr. 6
VI a3	Gr. 7	Gr. 8	Gr. 9
VI a4	Gr. 10	Gr. 11	Gr. 12

Nombre de groupes = 1 variable x 4 niveaux x 1 variable x 3 niveaux = 12

Ex. : la variable « tranche d'âge » a quatre niveaux : 21-25, 26-30, 31-35, 36-40 ans (de VI a1 à VI a4) ; la seconde variable « privation de sommeil » à 3 niveaux : 2, 4 et 6 heures de privation (de VI b1 à VI b3).

Une VI à deux niveaux, une VI à trois niveaux et une VI à un niveau

		VI c1	VI c2	VI c3	VI c4
VI a1	VI b1	Gr. 1	Gr. 2	Gr. 3	Gr. 4
	VI b2	Gr. 5	Gr. 6	Gr. 7	Gr. 8
	VI b3	Gr. 9	Gr. 10	Gr. 11	Gr. 12
VI a2	VI b1	Gr. 13	Gr. 14	Gr. 15	Gr. 16
	VI b2	Gr. 17	Gr. 18	Gr. 19	Gr. 20
	VI b3	Gr. 21	Gr. 22	Gr. 23	Gr. 24

Nombre de groupes = 1 variable x 2 niveaux x 1 variable x 3 niveaux x 1 variable x 4 niveaux = 24

Ex. : la variable « sexe » (VI a, 2 niveaux) ; la variable « tranche d'âge » (VI b, 3 niveaux) et la variable « privation de sommeil » (VI c, 4 niveaux).

Difficultés posées par les plans de recherche à groupes indépendants

Le problème principal est celui de l'équivalence des groupes. La question est de savoir si les effets observés sur les variables dépendantes dépendent uniquement des niveaux de(des) variable(s) indépendante(s). En effet, il est possible que d'autres variables, comme l'âge, le sexe, le niveau d'étude, etc., aient un impact sur la réponse des sujets.

Ces différences entre les groupes de sujets peuvent apparaître déjà au niveau même de la constitution des groupes (par exemple, vous recrutez les sujets d'un groupe en téléphonant chez eux et ceux de l'autre groupe en mettant une annonce dans un journal, au résultat vous n'obtiendrez pas tout à fait les

mêmes types de sujets). Vous devez donc, avant toute manipulation de la variable indépendante, former des groupes qui soient au plus semblables.

Plusieurs moyens sont possibles : vous pouvez chercher à regrouper des sujets qui ont des caractéristiques systématiquement identiques (par exemple, vous vous dites que pour éliminer l'influence du sexe, vous n'allez prendre que des hommes ou que des femmes). Cependant, en faisant ainsi, vous limitez la validité de vos résultats à un type de sujets. La seconde méthode consiste à faire appel au hasard pour sélectionner les sujets dans la population plus globale. Le hasard neutralise ainsi les biais qui pourraient résulter du fait de choisir vous-mêmes les sujets, même avec les meilleures intentions. Une troisième méthode consiste à apparier[4] les sujets deux par deux (par exemple, si vous avez un homme de quarante ans, niveau baccalauréat, marié avec enfants et au chômage dans un groupe, vous essaierez de trouver son homologue pour l'autre groupe, avec des caractéristiques les plus proches possible).

Une dernière méthode consiste à prendre un groupe homogène assez grand et de faire une sorte de prétest sur une variable de caractéristique ou sur une variable dépendante, puis de choisir le même nombre de sujets aux deux extrêmes de la distribution en rejetant ceux qui sont au milieu. Par exemple, vous voulez comparer le parcours de sevrage et après sevrage d'alcooliques selon leur niveau de dépression au départ. Vous allez faire passer une échelle de dépression à 100 sujets, puis vous ne retiendrez pour la suite de l'étude, que 20 de ceux qui sont le moins dépressifs (groupe faible dépression) et les 20 avec les plus hauts scores de dépression (groupe forte dépression). Vous obtenez ainsi deux groupes assez homogènes par ailleurs, mais qui sont très contrastés sur le plan de la variable intéressante : la dépression.

[4] Et non pas « appareiller », comme on peut le lire dans certains mémoires !

Bien entendu, toutes ces précautions de sélection et recrutement des sujets sont limitées par le nombre de sujets disponibles et d'autres contraintes. La constitution du groupe de contrôle (celui qui n'est pas soumis au niveau de la variable dont on veut mesurer les effets), par exemple, est souvent rendue très difficile par la nature même du groupe expérimental (celui qui va être soumis au niveau d'effet de la variable). Pour des toxicomanes en cure de sevrage, la variable indépendante étant « cure-pas cure », le groupe de contrôle est constitué de toxicomanes qui ne sont pas en cure. Mais, alors qu'il est facile de recruter les sujets en cure (on dit que c'est une population « captive », ils s'ennuient en centre de cure et attendent avec impatience de rencontrer l'étudiant en psychologie qui va les écouter et les distraire…), les toxicomanes chez eux, dans la rue, dans les associations d'aide, sont excessivement volatiles, ne viennent pas aux rendez-vous, n'ont que faire de répondre à vos questionnaires, etc. Sans parler du fait que certains peuvent avoir des traitements, des suivis psychologiques, qui font qu'ils ne sont pas « sans rien ».

À ces difficultés s'ajoute le fait, qu'idéalement, toute recherche devrait s'appuyer sur un nombre le plus important possible de sujets, car les résultats sont une sorte de « loi générale » que vous tirez de la répétition de l'expérience, de la mesure, et plus on répète et plus c'est valable. Cette nécessité du grand nombre entre souvent en conflit avec la nécessité d'éliminer des sujets pour atteindre à une équivalence des groupes. Aussi, on n'obtient jamais une équivalence parfaite, on ne fait qu'y tendre et vous verrez dans les parties recherche et statistique du manuel qu'il existe des moyens de tester le degré de cette équivalence, que « toutes choses sont égales par ailleurs ».

1.3.2. Plans de recherche à mesures répétées

Ici, c'est un seul et même groupe de sujets qui est confronté à tous les niveaux de la(les) variable(s) indépendante(s). Le premier niveau de la variable est souvent une situation de contrôle, une « ligne de base », à partir de laquelle les autres

niveaux seront évalués en fonction de cette situation de base. Souvent aussi, la mesure répétée concerne la variable « Temps » : ce qui change c'est la situation interne des sujets au cours du temps. Quelques exemples pour comprendre :

Une VI à deux niveaux (un test – un retest)

<table>
<tr><td>VI a1</td><td>Gr. 1^1</td><td>Test</td><td rowspan="2">Nombre de tests = 1 variable x 2 niveaux = 2

Ex. : la variable « avant / après cure de sevrage » (VI a, 2 niveaux). On mesure la dépression et l'anxiété (variables dépendantes), avant (Gr. 1^1, Test, situation de contrôle, ligne de base) et après 1 mois de cure de sevrage alcoolique (Gr. 1^2, Retest).</td></tr>
<tr><td>VI a2</td><td>Gr. 1^2</td><td>Retest</td></tr>
</table>

Une VI à cinq niveaux (un test + quatre retests)

<table>
<tr><td>VI a1</td><td>Gr. 1^1</td><td>Test</td><td rowspan="5">Nombre de tests = 1 variable x 5 niveaux = 5

Ex. : la variable « avant / après cure de sevrage à 1 mois / à 6 mois / à 12 mois / à 18 mois » (5 niveaux). On mesure la consommation d'alcool (variable dépendante), avant (VI a1, test, situation de contrôle, ligne de base) et après 1 mois de cure de sevrage alcoolique (VI a2, retest), puis à 6 mois (VI a3), 12 mois (VI a4) et 18 mois après la fin de la cure (retests).</td></tr>
<tr><td>VI a2</td><td>Gr. 1^2</td><td>Retest</td></tr>
<tr><td>VI a3</td><td>Gr. 1^3</td><td>Retest</td></tr>
<tr><td>VI a4</td><td>Gr. 1^4</td><td>Retest</td></tr>
<tr><td>VI a5</td><td>Gr. 1^5</td><td>Retest</td></tr>
</table>

Une VI à quatre niveaux + une VI à trois niveaux (12 tests)

<table>
<tr><th></th><th>VI b1</th><th>VI b2</th><th>VI b3</th><th rowspan="5">Nombre de tests = 1 variable x 3 niveaux x 1 variable x 4 niveaux = 12

Ex. : la variable « bruit ambiant » (3 niveaux : 1) pas de bruit (V b1) ; 2) faible bruit (V b2) ; 3) bruit fort (V b3)) est testée avec la variable « difficulté de la tâche » (4 niveaux : 1) facile (V a1) ; 2) un peu difficile (V a2) ; 3) difficile (V a3) ; 4) très difficile (V a4)). Ainsi, chaque sujet est soumis aux 12 situations qui combinent les différents niveaux des deux variables indépendantes. Un tel plan où la variable temps ne joue pas (les 12 expériences peuvent être menées dans la même matinée), peut concerner les deux exemples précédents, avec une seule variable à 2 ou 4 niveaux, testée dans un court laps de temps.</th></tr>
<tr><td>VI a1</td><td>Gr. 1^{1}</td><td>Gr. 1^{2}</td><td>Gr. 1^{3}</td></tr>
<tr><td>VI a2</td><td>Gr. 1^{4}</td><td>Gr. 1^{5}</td><td>Gr. 1^{6}</td></tr>
<tr><td>VI a3</td><td>Gr. 1^{7}</td><td>Gr. 1^{8}</td><td>Gr. 1^{9}</td></tr>
<tr><td>VI a4</td><td>Gr. 1^{10}</td><td>Gr. 1^{11}</td><td>Gr. 1^{12}</td></tr>
</table>

Difficultés posées par les plans de recherche à mesures répétées

Ici, le problème de l'équivalence des groupes de sujets semble disparaître, puisque nous n'avons qu'un seul groupe de sujets. Cependant, même si l'on n'a plus qu'un seul groupe, il est bon que les sujets de ce groupe ne soient pas trop disparates, en ce qui

concerne certaines de leurs caractéristiques reliées à la variable dépendante, il faut une sorte d'équivalence des sujets (chaque sujet avec son ensemble de test et retests forme bien une sorte de « groupe »). Par exemple, si vous étudiez l'anxiété chez des alcooliques au cours du temps (avant et après sevrage), il est préférable que ces alcooliques soient de façon prépondérante tous addictés à l'alcool et qu'il n'y ait pas parmi eux des addictés aux drogues ou de gros consommateurs de médicaments psychotropes. Car dans ce cas, le niveau d'anxiété ne dépendrait pas seulement du niveau de consommation alcoolique, mais aussi des autres consommations psychotropes qui ne sont pas, *a priori*, contrôlées par l'étude. Par contre, si vous avez à votre disposition un très grand nombre de sujets (par exemple, supérieur à 300), vous pouvez faire confiance au hasard pour homogénéiser l'ensemble des consommations toutes confondues et, à ce moment-là, après contrôle des consommations, faire une étude, non plus sur l'addiction alcoolique, mais l'addiction aux substances en général (en prévoyant peut-être un traitement statistique *a posteriori* sur des sous-groupes de consommateurs).

Un autre problème important de la répétition des mesures est le phénomène de **régression statistique**. Il s'agit, au fur et à mesure des mesures (si vous faites un simple test et retest, donc seulement deux mesures, le problème ne se pose pas), d'une tendance à ce que les scores des sujets extrêmes se resserrent autour de la moyenne. Pour y pallier, il suffit d'utiliser à l'envers la technique de la constitution de groupes extrêmes décrite à la page précédente (**Plans à groupes indépendants**) : ici nous ne voulons pas de sujets extrêmes, donc sur 100 sujets dépressifs, vous ne gardez pour l'étude que les 40 sujets au centre de la distribution des scores et vous rejetez les 30 moins dépressifs et les 30 plus dépressifs.

Une autre difficulté importante des plans à mesures répétées est l'**attrition** : la perte de sujets entre les mesures. Pour diverses raisons : médicales, professionnelles, familiales, de convenances personnelles, parfois tout simplement le décès, font qu'il n'est plus possible de rencontrer certains sujets 3 mois plus tard pour

un retest. Plus le temps passe entre deux mesures et plus il y a de mesures et plus le taux d'attrition est important. C'est pour cette raison qu'il vous faut, dès le départ, prévoir large, en comptant par exemple 5 % ou 10 % d'attrition. Si le nombre de sujets au départ n'est pas assez important, vous risquez de ne plus en avoir suffisamment pour valider l'étude, ou vous allez être confrontés à des problèmes d'hétérogénéité que le hasard et le grand nombre avaient gommés, mais qui ressurgissent avec la raréfaction des sujets.

Un autre problème est la **variabilité des caractéristiques des sujets dans le temps**. Au départ, vous avez fait un effort pour obtenir un groupe le moins disparate possible. Cependant, si les mesures se répètent sur des mois ou des années, cette homogénéité va peu à peu disparaître. Par exemple, vous étudiez la dépression à 3, 6, 12, 18 mois ; mais vos résultats seront faussés, car entre-temps certains sujets peuvent vivre des épreuves (séparation, deuil, chômage, accident) qui peuvent affecter leur score de dépression, par rapport aux sujets qui n'auront pas vécu ces épreuves. À l'inverse, des sujets peuvent naturellement « maturer », trouver un équilibre personnel, changer de vie (se marier, trouver un emploi, gagner un gain au loto…), ce qui va influencer leur dépression dans le bon sens. Là encore, seul le grand nombre peut gommer un peu ces disparités, ou permettre d'éliminer au fur et à mesure les sujets « non-conformes » sans que l'effectif devienne statistiquement inintéressant (car bien sûr, vous ne pouvez pas comparer les 100 sujets au départ, avec les 50 sujets à la fin, vous ne pouvez retenir que les 50 sujets qui ont pu être évalués, et au départ et à la fin).

Autre problème, lorsque la répétition des mesures s'inscrit au contraire sur le cours terme : la **démotivation des sujets**, qui peut résulter de la fatigue (physique, nerveuse, intellectuelle, parfois émotionnelle), de la lassitude, de l'agacement. Passer une matinée à repasser les mêmes tests peut effectivement devenir très pénible et cela va représenter une variable, difficile à contrôler, qui jouera sur les réponses des sujets. Cette démotivation peut d'ailleurs apparaître au cours d'un seul test, si

le questionnaire est trop long ou si l'on utilise trop de questionnaires et que le temps de passation dépasse l'heure. De plus, cette démotivation ne jouera pas de la même manière selon les sujets : un étudiant à l'esprit vif et habitué à lire et réfléchir se fatiguera moins vite qu'un sujet qui ne lit jamais, parce que ses capacités de lecture sont très faibles et lui demandent un grand effort (et dont le cerveau et les capacités seraient détériorés du fait de consommations psychotropes, par exemple…).

Pour contrer ce problème de la démotivation, plusieurs solutions, selon les cas : réduire la taille ou le nombre des questionnaires, des tests, aménager des pauses (avec petits gâteaux et café, mais attention aux interactions sociales entre sujets ou avec l'expérimentateur pendant les pauses ou à l'influence du café !), choisir des sujets de même performance (mais là, attention au biais écologique : vos résultats seront difficilement généralisables à des populations plus larges), vous pouvez aussi rémunérer vos sujets pour les motiver, à condition d'avoir le budget pour…

Un autre problème peut apparaître dans l'**effet d'ordre**. Dans le cadre d'une répétition à court terme, le fait de faire passer un test avant l'autre peut jouer sur la réponse au second. Cela peut être contré en alternant de façon aléatoire ou systématique l'ordre de passation des tests pour chaque sujet.

Dans le même registre, mais à long terme apparaît le **problème de l'habituation**. Les sujets s'habituent au test au fur et à mesure des passations (c'est le coup classique du vieux patient psychiatrique, qui sourit en coin, en voyant le jeune étudiant en psychologie débarquer avec son Rorschach ou son TAT), de sorte que les réponses reflètent davantage cette habitude que l'état du sujet lui-même. Certains tests comportent deux versions différentes, mais équivalentes, pour le test et le retest (comme l'*Eysenck Personality Inventory*, formulaires A et B). Sans cela, il n'y a guère de solution à ce problème.

La plupart des problèmes présentés ci-dessus peuvent être neutralisés par les plans combinés.

1.3.3. Plans de recherche combinés

Comme le nom l'indique, les plans combinés associent une comparaison entre groupes indépendants et une répétition de la mesure. En conséquence, un plan combiné fait appel à au moins deux variables indépendantes : une variable « Groupe » et une variable « Répétition ». Cependant, le second exemple ci-dessous, vous montre qu'un plus grand nombre de variables indépendantes peuvent être utilisées :

Une VI à deux niveaux (groupe) + une VI à deux niveaux (répétition)

	VI b1	**VI b2**	
VI a1	**Gr. 1^1**	**Gr. 2^1**	**Test**
VI a2	**Gr. 1^2**	**Gr. 2^2**	**Retest**

Nombre de groupes = 1 variable x 2 niveaux = 2
Nombre de tests = 1 variable x 2 niveaux x 1 variable x 2 niveaux = 4

Ex. : au cours d'une cure de sevrage alcoolique, la variable « psychothérapie » (VI b) à deux niveaux, définit deux groupes (Gr. 1 et Gr. 2), respectivement avec (VI b1) et sans (VI b2) psychothérapie. Les variables dépendantes, par exemple, anxiété et dépression, sont mesurées suivant la variable indépendante « Répétition » (VI a) : au début de la cure (VI a1) et après 3 mois de cure (VI a2), ce qui donne les deux tests du groupe 1 (Gr. 1^1 et Gr. 1^2) et les deux tests du groupe 2 (Gr. 2^1 et Gr. 2^2).

Un VI à quatre niveaux (critère groupe 1) + un VI à trois niveaux (critère groupe 2) + un VI à quatre niveaux (répétition)

		VI c1	**VI c2**	**VI c3**	**VI c4**
VI a1	**VI b1**	Gr. 1^1	Gr. 1^2	Gr. 1^3	Gr. 1^4
	VI b2	Gr. 2^1	Gr. 2^2	Gr. 2^3	Gr. 2^4
	VI b3	Gr. 3^1	Gr. 3^2	Gr. 3^3	Gr. 3^4
VI a2	**VI b4**	Gr. 4^1	Gr. 4^2	Gr. 4^3	Gr. 4^4
	VI b5	Gr. 5^1	Gr. 5^2	Gr. 5^3	Gr. 5^4
	VI b6	Gr. 6^1	Gr. 6^2	Gr. 6^3	Gr. 6^4
VI a3	**VI b7**	Gr. 7^1	Gr. 7^2	Gr. 7^3	Gr. 7^4
	VI b8	Gr. 8^1	Gr. 8^2	Gr. 8^3	Gr. 8^4
	VI b9	Gr. 9^1	Gr. 9^2	Gr. 9^3	Gr. 9^4
VI a4	**VI b10**	Gr. 10^1	Gr. 10^2	Gr. 10^3	Gr. 10^4
	VI b11	Gr. 11^1	Gr. 11^2	Gr. 11^3	Gr. 11^4
	VI b12	Gr. 12^1	Gr. 12^2	Gr. 12^3	Gr. 12^4
		Test	**Retest**	**Retest**	**Retest**

Nombre de groupes = 1 variable x 4 niveaux x 1 variable x 3 niveaux = 12
Nombre de tests = 1 variable x 4 niveaux x 1 variable x 3 niveaux x 1 variable x 4 niveaux = 48

Ex. : la variable indépendante « Méthode d'apprentissage » a quatre niveaux (VI a1 à VI a4), définissant quatre super-groupes, qui sont eux-mêmes subdivisés en trois sous-groupes selon la variable indépendante « Durée d'apprentissage » (VI b, par exemple, 10, 20 et 30 minutes), ce qui donne en tout douze groupes (de Gr. 1 à Gr. 12). De plus, ces douze groupes sont soumis à une variable indépendante de « Répétition » (VI c) de la mesure de la récupération mémorielle, quatre fois : à 1 heure (VI c1, test), 1 jour (VI c2, retest), 8 jours (VI c3, retest) et 15 jours (VI c4, dernier retest). Ce qui nous donne bien 12 x 4 = 48 tests pour ce protocole.

Les plans de recherche combinés cumulent les problèmes des plans de recherche par groupes indépendants et ceux à mesures répétées (voir les pages les concernant). Plus généralement, ces plans de recherche assez complexes (que l'on peut déconseiller aux étudiants) nécessitent un très grand nombre de sujets au départ, afin de garder une valeur statistique des résultats, malgré les inévitables attritions au fil du temps…

1.4. Les types de variables

Hormis la distinction fondamentale entre variables indépendantes (VI) et dépendantes (VD), il est important de connaître les différentes qualités attribuées aux variables, car de ces qualités peuvent dépendre le choix des recettes statistiques utilisées et certains choix méthodologiques. Les variables, au point de vue de leur nature, se rangent en deux catégories, les variables quantitatives et qualitatives :

1) Variables quantitatives

Les variables quantitatives (on peut dire aussi variables d'échelle – *scale* en anglais – ou continues) correspondent à une grandeur naturelle, une mesure. Elles sont aussi dites continues, car elles peuvent prendre toutes les valeurs possibles dans un intervalle.[5] Quelques exemples de variables quantitatives :

La température indiquée par un thermomètre entre 0° et 100° centigrades : 12°, 0,5°, 24,8°, 50°, etc.

L'âge de sujets entre 18 et 65 ans : 5, 10, 23, 35, 54, 78 ans, etc.

Les scores de dépression à une échelle de dépression entre 0 et 13 : 5, 8, 6, 10, 11, etc.

Les scores de performance à une épreuve de mémorisation entre 0 et 10 : 2, 7, 6, 5, 3, etc.

---

[5] En réalité, un « intervalle » apparaîtra si l'on ne considère que les valeurs entières (nombres entiers, ex. : 1, 12, 88), au lieu des valeurs avec décimales (nombres réels, ex. : 1,5, 12,52, 88,54089). Voir plus loin les variables nominales.

2) Variables qualitatives

Les variables qualitatives (on peut dire aussi variables discrètes, ou discontinues, ou catégorielles) ne concernent que certaines valeurs ou tranches de valeurs dans un intervalle ou encore des catégories ou classes de qualités. Quelques exemples de variables qualitatives :

Des classes d'âge : classe 1 entre 18-25 ans, classe 2 entre 26-35 ans, classe 3 entre 36-45 ans, classe 4 à partir de 46 ans et au-dessus.

Des niveaux de dépression : pas de dépression entre 0 et 3 ; dépression légère entre 4 et 7 ; dépression sérieuse entre 8 et 11 ; dépression grave à partir de 11 et au-dessus.

Les catégories de genre : « Homme » et « Femme ».

Les catégories socioprofessionnelles (CSP) : « Agriculteur », « Cadre, profession intellectuelle supérieure », « Profession intermédiaire », « Employé », « Ouvrier », « Retraité », etc.

Vous aurez remarqué que l'on peut subdiviser les variables qualitatives en deux types :

Les variables ordinales : il s'agit des classes de mesures, comme les tranches d'âge, les niveaux de température ou de dépression. Pour simplifier le traitement d'une variable numérique, ou pour la rendre plus parlante, on peut regrouper le caractère continu de la variable par tranches, qui sont réparties sur l'intervalle d'une façon ordonnée (par exemple, tranches de 0 à 5, de 6 à 10, de 11 à 15, de 16 à 20, etc.).

Les variables nominales : ne concernent en aucune façon des valeurs numériques, mais de véritables catégories, comme le fait d'être « Homme » ou « Femme », « Ouvrier » ou « Employé ».

Cependant, tout comme pour le choix de VI et de VD, il y a dans le choix de la qualité des variables une part de conventionnel : le chercheur décide librement si telle variable est prise à un niveau quantitatif ou qualitatif. Par exemple, si l'on reprend le cas de la variable continue des températures entre 0°

et 100°, cette variable est vraiment continue si on accepte toutes les unités de degrés entre 0 et 100 (12°, 15°, 5°, 78°, etc.). Dans certains cas, on peut aller plus loin et accepter toutes les mesures jusqu'aux fractions de degré, les millièmes (12,923°, 56,047°, etc.).

Pourtant, il est toujours possible, pour un chercheur, de décider qu'il mesure la température en millièmes de degrés et de découper l'intervalle entre 0° et 100° en 100 parties (ce ne sont plus des degrés, mais des noms de tranches), la variable devenant alors discontinue et ordinale (12 signifie entre 12,000 et 12,999 ; 13 signifie entre 13,000 et 13,999 ; et ainsi de suite).

Un autre choix très libre du chercheur réside dans le nom donné aux catégories d'une variable qualitative catégorielle. Cette liberté vous permet de transformer du texte en chiffre, en numéros (qui ne sont pas des mesures quantitatives évidemment). C'est ce qui arrive dans certains questionnaires ou inventaires, où le sujet doit répondre à l'aide d'une échelle du type : « Jamais », « Parfois », « Souvent », « Très souvent » ; ce que le chercheur traduit en un score entre « 0 » et « 3 ». Autres exemples : « Oui », « Non » deviendront « 1 » et « 0 » ; « Homme » et « Femme », « 1 » et « 2 », ou l'inverse, etc.

2. FAIRE UNE RECHERCHE

Les pages de ce chapitre du manuel présentent les différentes parties d'une recherche, qui sont aussi les différentes parties d'un mémoire ou d'un article scientifique (la différence entre mémoire et article scientifique tient le plus souvent à la taille du texte : dans un mémoire on peut davantage s'étaler et entrer dans certains détails ou reprendre des données antérieures pour en faire un exposé assez étendu, alors que l'article est généralement plus succinct. Cependant, cette notion de taille peut s'estomper quelque peu, si l'on considère que de plus en plus les enseignants (surbookés) exigent des mémoires « courts » et que certaines revues scientifiques acceptent des articles de fond et de haut vol plus ou moins importants en taille).

2.1. L'introduction

L'introduction à votre étude doit commencer par des considérations assez générales, qui évoquent la problématique sur laquelle vous travaillez, et qui justifient en quelque sorte l'intérêt de votre étude. Pour introduire ainsi votre sujet, il est très important de référencer chacune de vos affirmations, informations. Il est aussi important de commencer à donner des données chiffrées.

Exemple :

Ne pas écrire : « En général, le taux de rechute après un traitement est très important. »

Mais écrire : « En général, le taux de rechute après un traitement varie entre 50 % et 90 % (Browell, Marlatt, Lichtenstein, & Wilson, 1986). »

Vous donnez ainsi une estimation chiffrée du taux de rechute et cette estimation renvoie à une étude validée et publiée dans une revue scientifique officielle (renoncez aux citations issues de magazines de vulgarisation, ou pire de magazines tout court ! – même si c'est le docteur Machin qui l'a dit). En donnant ainsi des chiffres et en les référençant, vous donnez au lecteur des informations précises et fiables qui vont l'éclairer (et qu'il peut retrouver lui-même et vérifier) et vous lui montrez que vous connaissez bien votre sujet, pour avoir lu (ou connu) les articles en question.

Nota : en ce qui concerne le « avoir lu ». Bien sûr, dans l'idéal, il est toujours bon et préférable d'avoir lu les articles que l'on cite. Cependant, dans la pratique, l'étudiant, comme le chercheur confirmé, se heurtent à des limites de temps disponible pour lire, à l'énergie, parfois à la motivation, nécessaires. Il existe aussi en science un souci d'économie de l'effort que l'on peut formuler ainsi : quelqu'un a déjà fait l'effort de lire un article et le cite avec un commentaire suffisant pour en prendre la pleine mesure, donc pourquoi produire une redondance inutile de l'effort dans la communauté scientifique et ne pas me contenter de reprendre la citation du précédent auteur. C'est ce qui se passe, en fait, pour la plupart des citations : on se contente de reprendre la citation de l'autre. Par exemple, vous n'avez jamais lu l'article de Browell, Marlatt, Lichtenstein et Wilson (1986) concernant les taux de rechute alcoolique, mais vous avez repris cette information dans un autre article. Parfois, l'honnêteté entraîne certains à dire, entre parenthèses (Dupont, 1985, cité dans Durand, 1999), mais

le procédé ne fait qu'alourdir le texte et accroître artificiellement la bibliographie.

Il est vrai que la citation de la citation ressemble un peu à la blague de « l'homme qu'a vu l'homme qu'a vu l'homme (...) qu'a vu la bête » et cela peut conduire à des effets pervers, dignes des déformations qui apparaissent dans une rumeur. Un exemple frappant, le fameux article de Goodman (1990) qui définit les critères de l'addiction. Rares sont les personnes qui ont vraiment lu cet article (qui plus est, en anglais), mais par contre, nombreuses sont celles qui y font référence et qui utilisent sans réfléchir les critères de Goodman pour construire un instrument d'évaluation ou sélectionner leurs sujets. La lecture que nous avons faite de cet « article » a pu montrer qu'il ne s'agissait que d'un *commentary* (bien qu'il soit publié dans une revue sérieuse), à caractère simplement spéculatif, et non d'une véritable démonstration des critères. Par contre, l'article explicite des implications pratiques et théoriques très importantes, qui sont ignorées, faute d'avoir été lues.

En tout état de cause, il faut s'efforcer au fil des années de recherche, de lire au moins les articles clés (et souvent célèbres), afin de savoir exactement, et de première main, de quoi ça parle.

Exemples d'introduction :

Dans de nombreux pays, l'alcoolisme est un sérieux problème de santé publique. Ce problème est aggravé du fait que les traitements sont trop souvent suivis de rechute. En général, le taux de rechute après un traitement varie entre 50 % et 90 % (Browell, Marlatt, Lichtenstein, & Wilson, 1986). D'après Chiauzzi (1989), entre 66 % et 75 % des abuseurs de drogues ou d'alcool connaîtront une rechute dans les trois mois suivant le traitement. Armor, Polich et Stambul (1978) indiquent que seulement 10 % des alcooliques traités maintiennent leur abstinence plus de 2 ans. Dans une étude de suivi sur 2 ans au Chili, après un traitement en centre de réhabilitation, 63 % des

individus retournent à l'alcool (Bunout, Maza, Hirsch, Petermann, Godoy, & Pérez, 1997).

L'abandon du traitement, qui entraîne ou est la conséquence de la rechute alcoolique, présente des taux semblables aux rechutes directes. Baekeland et Lundwall (1975) rapportent des taux d'abandon précoce de 52 % à 75 %. Une étude de Laflamme-Cusson et Moisan (1993) fait ressortir des pourcentages comparables au Canada : 25 % des patients abandonnent leur demande d'aide après l'entrevue d'évaluation, 50 % poursuivent pendant plus de quatre semaines et seulement 15 % persistent plus de 6 mois. Dans une autre étude conduite au Canada (Brochu, Landry, Bergeron, & Chiocchio, 1995), des taux d'abandon semblables sont observés aux différentes phases de traitement : 17 % des patients abandonnent leur demande d'aide après l'entrevue d'évaluation, 45 % abandonnent avant la deuxième phase de traitement et seulement 12 % persistent au-delà de six mois.

Toujours dans l'introduction, il ne faut pas vous contenter d'une seule donnée et d'une seule référence. Vous devez avoir suffisamment exploré la littérature scientifique (y compris et surtout internationale) pour pouvoir présenter plusieurs données référencées, qui sont souvent contradictoires ou sensiblement différentes. Si vous cherchez bien, peut-être pourrez-vous utiliser une revue de la littérature avec une méta-analyse[6] : un gros article scientifique dans lequel le chercheur (ou une équipe) a fait l'effort de lire systématiquement l'ensemble des travaux portant sur un thème donné et en a tiré des conclusions intéressantes (pour vous).

[6] Une méta-analyse est une démarche scientifique systématique combinant les résultats d'une série d'études indépendantes sur un problème donné, selon un protocole reproductible. La méta-analyse permet une analyse plus précise des données par l'augmentation du nombre de cas étudiés et de tirer une conclusion globale.

Vous devez donc déjà, pouvoir esquisser ici une sorte de mini débat. Car, à la fin de la conclusion (de l'introduction), vous devez justifier l'intérêt de votre étude : parce que des données sont contradictoires, ou imprécises, ou mal produites, parce qu'il y a débat, discussion dans la communauté scientifique, alors, votre recherche est utile pour… il est utile, nécessaire de… Voilà par quoi une conclusion d'introduction devrait se terminer.

Exemple de conclusion à l'introduction :

Néanmoins, l'interprétation de l'ensemble des résultats de ces études ne permet pas d'arriver à des conclusions fermes et, bien qu'un certain nombre de variables pertinentes aient pu être identifiées, la principale critique qui peut être adressée à ces recherches porte sur l'absence de modèles théoriques et intégratifs dans l'étude de ce phénomène. Aussi, il est nécessaire d'utiliser un tel modèle intégratif et d'en faire la démonstration.

2.2. La problématique

Un principe fondamental de la recherche scientifique en psychologie peut se formuler ainsi : **« voir large, faire étroit »** (ou le *principe de l'entonnoir*, Fig. 2).

Cela signifie que vous devez d'abord ratisser large, puis, réduire votre problématique et votre modèle à quelque chose de plus étroit. Cette réduction, effet d'entonnoir, est nécessaire pour deux raisons : d'abord, vous ne pouvez pas traiter de toute la complexité du réel et vous serez amené à faire des choix (choisir un modèle parmi les modèles, choisir une population dans la population générale, choisir vos hypothèses, choisir vos variables – VI et VD).

La seconde raison est en continuité avec la première, c'est celle de la faisabilité : non seulement vous ne pouvez prétendre appréhender la complexité du réel humain (même si vous êtes une équipe de chercheurs), mais en plus, votre énergie, vos

connaissances, vos moyens matériels, financiers, institutionnels, socio-économiques, voire politiques, sont limités.



Figure 2 : Le principe de l'entonnoir dans la recherche.

Votre temps et vos objectifs sont aussi limités : on ne vous demande pas d'écrire une somme en 20 volumes sur 10 ans, on vous demande de faire une petite recherche en quelques mois et d'écrire un petit mémoire, un article, ou au pire une grosse thèse.[7]

---

[7] Il y eut une époque (au 20e siècle) où réaliser une thèse en 10 ans était la norme. Mais devant l'inflation des travaux scientifiques et l'avalanche permanente de nouveaux résultats, cette élaboration sur le long terme est devenue impossible, la thèse deviendrait interminable puisqu'il faudrait pratiquement chaque année la reprendre à la lumière des découvertes les plus récentes. Aussi, une thèse se déroule-t-elle normalement en 3 ans : une année pour préparer son sujet, explorer la littérature, repérer le terrain d'étude et établir la faisabilité ; une année

Dans la problématique, vous devez survoler les différents modèles qui ont tenté d'expliciter votre problématique. Puis, vous devez faire un choix (il est arbitraire, vous êtes libre, ce qui ne doit pas vous empêcher de justifier rationnellement votre choix), parmi les modèles, vous emparer de l'un d'eux comme base pour votre étude. C'est le *modèle de référence*.

Ensuite, vous devez vraiment montrer au lecteur à quel point vous vous êtes emparé du modèle de référence en formulant votre propre *mini-modèle*, qui est une forme simplifiée (pas forcément) et adaptée du modèle de référence, à la problématique concrète à laquelle vous voulez apporter une réponse. Par exemple, si vous faites le choix d'aborder la problématique de la rechute addictive au travers du modèle de référence de la théorie du renversement (Apter), vous pouvez réduire cette théorie complexe à certains concepts et indicateurs (par exemple, la paire d'états métamotivationnels télique-paratélique) et imaginer (c'est là votre mini-modèle) des éléments de processus explicatifs de la rechute à partir de ces concepts. Vous ferez cela, tout en le reliant à votre terrain de recherche (par exemple, des toxicomanes en postcure).

Bien entendu, au passage, votre problématique et votre mini-modèle peuvent être étayés par des travaux antérieurs éventuels que vous devez citer. Par exemple, vous pouvez en toute modestie reprendre un protocole expérimental (les fameuses études réplicatives – *replication studies* –, l'un des grands principes de la recherche scientifique étant la reproductibilité, vous y contribuerez) et citer précisément les résultats antérieurs à votre propre étude.

2.3. Les hypothèses

La recherche scientifique consiste à faire des prédictions… tout comme la voyante cartomancienne vous prédit l'avenir, le turfiste les chevaux gagnants de la prochaine course ou le

---

pour la réalisation du protocole de recherche sur le terrain ; une année pour traiter les données, la rédaction et la soutenance. *Not more !*

météorologue le temps qu'il fera ce week-end. Cependant, les prédictions du chercheur sont tout à fait différentes. Tout d'abord, au contraire de la tireuse de cartes, nos prédictions s'appliquent sur un modèle plausible et rationnel (on ne peut pas dire que la correspondance entre le tirage aléatoire de cartes à jouer et des événements futurs représente un « modèle » plausible et rationnel, sauf à le croire…).

Ensuite, nos prédictions n'ont pas une valeur pragmatique, comme le bulletin météo, mais théorique : nous ne gagnons rien de concret dans le fait que nos prédictions soient exactes ou non, ni pour nous-mêmes (le chercheur, s'il est rémunéré, touchera son salaire, que ses hypothèses soient vérifiées ou non), ni pour la société dans le court terme (personne n'attend le résultat de nos prédictions pour le lendemain ou le week-end prochain, par contre, la société est effectivement intéressée par nos résultats pour une action, une application, inscrites dans le long terme, par exemple, changer l'approche sanitaire de la problématique de la délinquance juvénile).

Nos prédictions scientifiques sont plausibles, s'appliquent sur un modèle étroit et rationnel et de plus, elles n'ont pas d'orientation ! Nous voulons dire « pas d'orientation autre qu'arbitraire ». Cela signifie que vous pouvez formuler vos hypothèses en positif ou en négatif. Par exemple, vous êtes libres de poser :

La cure de sevrage avec accompagnement psychologique aura un effet d'amélioration sur l'état psychologique (dépression et anxiété) par rapport au groupe de contrôle qui suit une cure de sevrage sans accompagnement psychologique.

Ou le contraire :

La cure de sevrage avec accompagnement psychologique n'améliorera pas l'état psychologique (dépression et anxiété) tout comme le groupe de contrôle qui suit une cure de sevrage sans accompagnement psychologique.

De plus, pour le chercheur, ses résultats sont, nous l'avons dit, pragmatiquement neutres, ce qui signifie que si une hypothèse est validée, c'est un résultat ; mais si une hypothèse n'est pas validée, c'est **aussi** un résultat ! Pour le chercheur, que l'accompagnement psychologique dans la cure de sevrage améliore ou non l'état psychologique des sujets est, dans les deux cas, un résultat intéressant : s'il y a amélioration, alors il ajoute une pierre au modèle qui dit que « l'accompagnement psychologique en cours de cure de sevrage est cliniquement intéressant », et s'il n'y a pas amélioration, il n'ajoute pas de pierre, voire, il en retire une et il faut alors envisager autre chose que l'accompagnement psychologique au cours du sevrage.

Cette liberté d'orientation, quant à la formulation des prédictions, fait que le chercheur, entre l'arrivée des résultats et la publication, est libre de choisir dans quel sens il publiera ses résultats :

Sens 1 : le chercheur présente les modèles de sa problématique, choisit parmi eux son modèle de référence, en tire son mini-modèle, à partir duquel il construit des hypothèses « classiques » ou « orthodoxes », c'est-à-dire qui vont dans le sens communément attendu. Et, dans le cas où les prédictions ne sont pas vérifiées, le chercheur proposera à la suite de la discussion de ses résultats un modèle « alternatif », « hétérodoxe », « révolutionnaire » (au sens de *révolution scientifique*).[8]

Sens 2 : le chercheur présente les modèles de sa problématique, mais se démarque d'emblée de ces modèles et propose, arguments scientifiques à l'appui, un mini-modèle « alternatif », « hétérodoxe », « révolutionnaire » (au sens de *révolution scientifique*),[10] à partir duquel il construit des hypothèses « alternatives », « hétérodoxes », « révolutionnaires », qui vont à contre-courant des hypothèses « classiques » ou « orthodoxes ». Et comme, dans ce cas, les prédictions du

[8] Kuhn, T. S. (1972). *La structure des révolutions scientifiques*. Trad. Française, Paris : Flammarion.

chercheur sont validées, dans la discussion des résultats, il confortera son modèle révolutionnaire et tâchera de lui apporter quelques réserves et critiques en toute modestie (de toute façon d'autres se chargeront bien de le faire, croyez-moi !).

L'une et l'autre orientation sont scientifiquement indifférentes, car comme on l'a dit, un résultat est toujours un résultat, quel que soit son orientation (qu'il valide ou pas l'hypothèse). Par contre, apparaissent ici des notions d'esthétique dans la présentation des travaux et de satisfaction narcissique du chercheur lui-même : dans le sens « 1 », le chercheur sacrifie son plaisir narcissique (si ses hypothèses ne sont pas validées) au profit d'une présentation classique ; dans le sens « 2 », le chercheur se fait plaisir (*Ouah ! qu'est-ce que je suis fort, mes prédictions à moi se sont réalisées !*), au profit d'une présentation avantageuse. Entre être classique, révolutionnaire ou chercheur puéril, c'est à chacun de se faire son opinion. Mais il est temps de parler de ces fameuses hypothèses.

Qu'est-ce qu'une hypothèse ?

Une hypothèse (à cet endroit, il s'agit d'hypothèse opérationnelle, voir le chapitre « **Les V.I. et les V.D.** » pour les différents niveaux d'hypothèses : général, de travail, opérationnel et statistique) est une prédiction (issue d'un modèle plausible et rationnel) qui pose une relation d'influence entre une(des) variable(s) indépendante(s) et une(des) variable(s) dépendante(s) (les niveaux de la VI ont un effet sur – sont explicatifs de – la valeur de la VD). Par exemple, dans l'hypothèse ci-dessous :

La cure de sevrage avec accompagnement psychologique (VI) aura un effet d'amélioration sur l'état psychologique (VD = dépression et anxiété) par rapport au groupe de contrôle qui suit une cure de sevrage sans accompagnement psychologique.

La VI est l'accompagnement psychologique, elle possède deux niveaux : avec (groupe expérimental) et sans (groupe de contrôle, on dit aussi témoin) accompagnement psychologique.

La VD est l'état psychologique, opérationnalisé comme scores à des échelles de dépression et d'anxiété.

L'influence est prédite de VI sur VD : l'accompagnement psychologique influence l'état psychologique.

Enfin, rappelez-vous au chapitre sur « **Les V.I. et les V.D.** », que le sens de la prédiction, de l'influence, entre ce qui est VI ou VD, vous appartient : ici le chercheur étudie l'influence de l'accompagnement psychologique sur l'état psychologique, mais un autre chercheur pourrait étudier l'influence de l'état psychologique sur l'accompagnement psychologique (par exemple, avec l'hypothèse plausible de ce que plus l'état psychologique est grave – anxiété et dépression – et plus l'accompagnement psychologique est inefficace du fait de la non-compliance[9] des sujets).

La VI est bien ce que manipule le chercheur (il divise son échantillon de sujets en deux groupes : ceux qui reçoivent et ceux qui ne reçoivent pas l'accompagnement psychologique) et la VD est bien la réponse des sujets aux questionnaires (dépression et anxiété). À ce propos, notez bien cette subtilité de langage : quand vous faites une prédiction sur une amélioration d'état psychologique, sur un changement de comportement ou d'attitude (sauf, dans ce cas, à pouvoir observer vous-mêmes le comportement), c'est par commodité langagière, simplification, que l'on dit dans l'hypothèse que les niveaux de la VI ont une influence sur l'état ou le comportement ; car en réalité, la prédiction se borne à porter sur la réponse aux instruments de

[9] Terme qui vient de l'anglais et qui est de plus en plus utilisé en français faute d'un mot équivalent et qui signifie : l'adhésion et la conformité d'un sujet à un protocole de soin ou d'étude (par exemple, respecter les rendez-vous, le cadre du travail, répondre aux consignes, prendre le traitement et respecter la posologie, etc.).

mesure (questionnaires, inventaires, tests). Pour ce qui concerne le lien prédictif entre comportements, attitudes, états psychologiques réels et la réponse aux instruments, c'est une autre histoire (cela dépend évidemment du degré de validité de l'instrument de mesure).

Rappel : voir le chapitre « **Les V.I. et les V.D.** » pour connaître les différents niveaux d'hypothèses : général, de travail, opérationnel et statistique.

Les types d'hypothèses selon les types de recherche

Au chapitre présentant « **Les types de recherche** », nous avions vu les quatre niveaux de la recherche : empirique, théorique, (quasi) expérimental et clinique. La possibilité de formuler des prédictions hypothético-déductives varie selon ces niveaux de la recherche.

- **Recherche empirique :** ces recherches se situent en amont des études expérimentales, souvent dans une visée exploratoire. Il s'agit d'explorer un nouveau terrain d'étude, de connaître une population, d'acquérir des indices et indicateurs de base. Dans ce cas, le chercheur peut ne faire que de l'observation systématique (la description détaillée de tout ce qu'il observe) et ne formuler qu'une simple hypothèse sur ce qu'il s'attend à observer (intensité, fréquences, durées d'un comportement, par exemple). Le chercheur peut aller un peu plus loin au plan empirique en portant des prédictions corrélationnelles entre deux variables et ses hypothèses corrélationnelles, dans une visée simplement inductive, vont alors prédire une covariation entre ces deux variables (par exemple, entre l'âge et les comportements d'imitation). Dans tous les cas, on ne peut pas vraiment parler de « manipulation » de variable indépendante par le chercheur et donc, de relation causale postulée dans l'hypothèse.
- **Recherche théorique :** la recherche théorique, souvent placée entre recherche empirique et recherche expérimentale, n'est pas destinée à formuler des hypothèses à tester

immédiatement. Il s'agit d'une étape au cours de laquelle, à partir des résultats des travaux empiriques (l'observation), un modèle est construit, dans sa confrontation aux autres modèles déjà existants. Ce modèle prévoit évidemment des processus, qui donnent lieu à hypothèses, mais tout ce travail de pionnier se borne à des hypothèses très générales (non encore opérationnelles), destinées à définir les grands axes d'un programme de recherche à venir.[10]

- **Recherche (quasi) expérimentale :** c'est dans ce type de recherche que l'hypothèse prend pleinement son essor, car nous sommes bien dans une visée hypothético-déductive, pour laquelle le chercheur prédit les conséquences causales de ses manipulations, selon le mini-modèle qu'il a choisi. Dans ce cas, il y a des VI (manipulées) et des VD (en réponse de la part des sujets) et l'hypothèse prédit l'obtention d'une relation précise, opérationnelle, entre les variations induites (VI) et les variations résultantes (VD).
- **Recherche clinique :** ici, la recherche est du même type que la recherche empirique, sinon que l'exploration ne porte que sur un champ restreint, fouillé en profondeur. Les seules hypothèses formulables porteront sur ce que le chercheur s'attend à observer chez un ou quelques individus, cela en regard des observations antérieures et de son modèle de référence. Dès que le chercheur commence à « manipuler » quelque chose, s'attendant alors à un résultat, il entre dans le champ expérimental. Dans ce cas, la portée de ce que le chercheur prédit et produit et reproduit ne pourra dépasser le sujet qu'il étudie (généralisation des résultats) que s'il se

---

[10] À ce niveau théorique, on peut rencontrer des hypothèses et leur test dans ce que l'on appelle des « expériences de pensée » (imaginaires) : vous décrivez d'une façon très plausible et rationnelle une véritable expérimentation scientifique et vous prédisez ses résultats (Einstein était coutumier du fait). Les simulations par ordinateur font partie de ce type de recherches théoriques.

tourne vers un échantillon de population, de fait, il ne s'agit plus de recherche clinique au sens où nous l'entendons ici.

Dans votre article ou votre mémoire, c'est à ce stade que vous reliez opérationnellement votre mini-modèle à la problématique et que vous faites l'exposé descriptif général de votre plan de recherche (les manipulations que vous allez faire et les résultats que vous en attendez). En conclusion, vous présentez vos hypothèses (voir le chapitre : « **Variable dépendantes et variables indépendantes** »).

2.4. La méthode

Dans cette partie de votre travail, vous devez exposer exactement la méthode (et non la « méthodologie », qui est l'ensemble des méthodes scientifiques) que vous avez suivie pour réaliser votre étude (on doit savoir ce que vous avez fait et on doit pouvoir reproduire votre méthode). Cet exposé comporte les éléments suivants :

Descriptif de votre population : il s'agit d'indiquer où et comment la population d'étude a été recrutée, puis de la décrire (nombre d'hommes, de femmes, nombre total ; moyenne, écart type, minimum et maximum des âges). Vous devez dire exactement comment vous avez rencontré les sujets (dans une institution, quel type ; par une petite annonce, spontanément dans un lieu public, à l'université, par téléphone, par effet boule de neige – une personne vous donne les coordonnées d'une ou plusieurs autre(s) qu'elle connaît et ainsi de suite –, etc.). Vous pouvez utiliser des tableaux chiffrés pour présenter certaines données. Notamment, lorsque vous présenterez les données démographiques. Précisez si les personnes ont rempli une Feuille de consentement. Précisez aussi vos critères d'inclusion et de non-inclusion dans la recherche.

Descriptif de votre protocole : vous devez expliquer clairement comment vous avez procédé : pour répartir vos échantillons dans les groupes (les critères), dans quelles conditions vous avez assuré vos variables indépendantes

(dispositif expérimental, lieu, conditions, durée, contrôle des variables parasites) ; quant, où, comment, vous avez enregistré vos variables dépendantes (le test, le retest ou le calendrier ou programme de répétition des mesures, etc.).

Descriptif des instruments de mesure : l'instrument qui établit les critères de sélection entre groupe expérimental et groupe témoin (la VI) et les instruments qui permettent de mesurer les variables dépendantes (VD), en donnant à chaque fois la référence la plus récente de validation de l'instrument, dans la langue des sujets. Dites si vous avez choisi de prendre en compte les données normatives (par exemple, à partir de quel seuil vous classez les sujets comme dépressifs ou non).

Bien entendu, vous êtes libre d'adapter le choix des informations et leur ordre de présentation selon le niveau de votre recherche et ses caractéristiques propres.

2.5. Les résultats

Les résultats sont présentés en trois parties :

1. Le **contrôle des variables descriptives (organismiques, démographiques) sur l'équivalence des groupes** afin d'assurer le lecteur que vous avez pris en compte le principe du « toutes choses égales par ailleurs ». Pour cela, vous donnez des tableaux chiffrés, qui précisent les résultats de vos tests statistiques sur ces variables (voir la partie « **L'équivalence des groupes** » du manuel). Normalement, vos tests ne doivent pas indiquer une différence significative des groupes au point de vue des variables descriptives. Dans le cas contraire, il faudra penser à apporter des réserves à vos résultats dans la discussion finale.

2. Le **contrôle des variables descriptives (organismiques, démographiques) sur les variables dépendantes** afin d'assurer le lecteur que vous avez pris en compte le principe du « toutes choses égales par ailleurs ». Pour cela, vous donnez des tableaux chiffrés, qui précisent les résultats de vos tests statistiques sur ces variables (voir la partie « **L'effet des**

variables descriptives » du manuel). Normalement, vos tests ne doivent pas indiquer une influence significative des variables descriptives sur vos résultats (les VD). Dans le cas contraire, il faudra penser à apporter des réserves à vos résultats dans la discussion finale.

3. La **présentation des résultats proprement dits**. Elle se fait sous la forme de tableaux chiffrés et commentés. Les commentaires sont brefs et ne dépassent pas la description des points saillants à relever dans le tableau (voir la partie « **Les statistiques inférentielles** » du manuel). Pour aller plus loin, attendez la partie suivante, la discussion des résultats.

Dans le chapitre « **Préparation de SPSS** » du manuel, vous verrez comment respecter les normes de présentation des tableaux de résultats.

2.6. La discussion

La discussion finale consiste à analyser vos résultats en regard de vos hypothèses, en partant du détail, pour aller progressivement vers le général. Pour chacune des hypothèses, analysez dans quelle mesure elle est ou non vérifiée. Analysez aussi l'éventuelle influence des variables descriptives, ainsi que l'influence de certaines conditions expérimentales, des limites d'effectif de l'échantillon, la disparité entre les groupes, l'attrition entre test et retest, etc.

Ensuite, dans la mesure où vos hypothèses ont été plus ou moins validées, vous devez expliquer comment cela renforce le modèle. Dans le cas contraire, vous devez trouver un équilibre entre ce qui limite la validité de vos résultats et ce qui remet en question le modèle : si vos résultats sont excessivement solides (gros échantillon, protocole impeccable), le modèle peut être remis en question ; si ce n'est pas le cas, vous vous contentez de porter des doutes sur le modèle, en toute modestie et vous faites appel au souhait de recherches ultérieures, pour y voir plus clair.

Vous pouvez aussi aller plus loin et commencer à suggérer des modifications au modèle, plus ou moins consistantes, ou

carrément vous suggérez un nouveau modèle et en produisez quelques éléments. Vous assortissez le tout d'hypothèses générales ou de travail, qui vont inspirer et esquisser les recherches futures (même si ce n'est pas vous qui entreprendrez de telles recherches).

2.7. La conclusion

Dans la conclusion de votre mémoire, ou de votre article, vous devez rappeler et résumer rapidement vos résultats et les replacer dans la perspective de la problématique initialement traitée. Montrer comment vos résultats éclairent ou n'éclairent pas cette problématique et quelles applications on pourrait ou pas envisager (une conclusion est généralement courte, ce paragraphe en apporte la preuve…).

2.8. Les références

À la fin de votre texte (mémoire ou article), vous devez lister vos références bibliographiques. **Évitez absolument les références orphelines !** Celles citées dans le texte et que vous oubliez de lister à la fin, ou encore celles que vous mettez dans la liste finale, pour gonfler artificiellement votre bibliographie, sans qu'elles correspondent à une citation dans le texte. Si vous pensez que des références orphelines de ce type sont utiles à lister pour le lecteur, vous pouvez créer une seconde liste de références intitulée « Bibliographie supplémentaire » ou « Pour en savoir plus » (cette seconde liste est acceptable dans un mémoire ou mieux dans un ouvrage, mais elle sera refusée dans un article).

Les références bibliographiques en liste sont classées selon trois méthodes possibles :

1) Alphabétique auteur : c'est la classification la plus commune et classique, c'est aussi la plus simple à réaliser. Les références dans le texte sont données avec nom d'auteur suivi de l'année et, dans la liste, les références sont classées par ordre alphabétique des auteurs.

Exemple, méthode 1 :

Dans le texte…

Jusqu'à présent l'œuvre de Stanton Peele a été présentée dans la littérature française comme « éclectique et descriptive, empruntant au constructivisme social et recourant à des concepts comportementaux et cognitifs » (Pedinielli, Rouan & Bertagne, 1997). (…). Les débats des spécialistes à propos de la toxicomanie, aux États-Unis comme en Europe (MacCoun & Reuter, 1997 ; Henrion, et al., 1995 ; Solal, 1997 ; Lucas, 1996) et les très récents rapports ministériels (Roques, 1998 ; Maestracci, 1999), montrent que les ouvrages de Peele (1975, 1985) et l'ensemble de son œuvre, n'ont pas encore perdu une once d'actualité et qu'il peut être utile de les examiner à nouveau.[11]

Dans la liste bibliographique finale…

1. Henrion, R., et al. (1995). Rapport de la commission de réflexion sur la drogue et la toxicomanie. Paris : La Documentation Française.
2. Lucas, B. (1996). Politique française en matière de drogue, la singularité du cas lyonnais. *Psychotropes*, 2, 75-95.
3. MacCoun, R. & Reuter P. (1997). Interpreting dutch cannabis policy: Reasoning by analogy in the legalization debate. *Science*, 278, 47-52.
4. Maestracci, N. (1999). Lutte contre la drogue et la toxicomanie. Rapport remis au Premier ministre Lionel Jospin, janvier, confidentiel.
5. Pedinielli, J.-L., Rouan G., & Bertagne, P. (1997). *Psychopathologie des addictions*. Paris : Presses Universitaires de France, Nodules.
6. Peele, S. & Brodsky, A. (1975). *Love and addiction*. New York: Taplinger.

[11] Loonis E. (2014). *Théorie générale de l'addiction : Introduction à l'hédonologie*. Éditions Éric Loonis.

7. Peele, S. (1985). *The meaning of addiction*. Lexington: Lexington Press.
8. Roques, B. (1998). *La dangerosité des drogues*. Paris : Odile Jacob, La Documentation Française.
9. Solal, J. F. (1997). Abstinence et substitution : histoire et actualité des traitements aux toxicomanes. *Pratiques Psychologiques*, 4, 13-24.

2) Alphabétique numérotée : ici, les auteurs sont classés dans la liste comme précédemment, dans l'ordre alphabétique et vous numérotez chaque référence dans la liste (sous Word avec une puce de numérotation). Cependant, dans le texte, au lieu de mettre nom d'auteur et année de publication, vous mettez le numéro de liste bibliographique. Problème : il faut d'abord faire comme pour la méthode 1, puis, si vous êtes certain d'avoir listé toutes vos références, remplacez dans le texte les (nom + année) par le numéro. Inconvénient : le procédé est source d'erreur et fait perdre du temps inutilement, de plus, si vous souhaitez ajouter une référence, votre numérotation est complètement décalée et à reprendre et, là encore, le risque d'erreur est grand.

Exemple, méthode 2 :

Dans le texte…

Jusqu'à présent l'œuvre de Stanton Peele a été présentée dans la littérature française comme « éclectique et descriptive, empruntant au constructivisme social et recourant à des concepts comportementaux et cognitifs » (5). (…). Les débats des spécialistes à propos de la toxicomanie, aux États-Unis comme en Europe (1, 2, 3, 9) et les très récents rapports ministériels (4, 8), montrent que les ouvrages de Peele (6, 7) et l'ensemble de son œuvre, n'ont pas encore perdu une once d'actualité et qu'il peut être utile de les examiner à nouveau.[12]

[12] Ibidem.

Dans la liste bibliographique finale…

1. Henrion, R., et al. (1995). Rapport de la commission de réflexion sur la drogue et la toxicomanie. Paris : La Documentation Française.
2. Lucas, B. (1996). Politique française en matière de drogue, la singularité du cas lyonnais. *Psychotropes*, 2, 75-95.
3. MacCoun, R. & Reuter P. (1997). Interpreting dutch cannabis policy: Reasoning by analogy in the legalization debate. *Science*, 278, 47-52.
4. Maestracci, N. (1999). Lutte contre la drogue et la toxicomanie. Rapport remis au Premier ministre Lionel Jospin, janvier, confidentiel.
5. Pedinielli, J.-L., Rouan G., & Bertagne, P. (1997). *Psychopathologie des addictions*. Paris : Presses Universitaires de France, Nodules.
6. Peele, S. & Brodsky, A. (1975). *Love and addiction*. New York: Taplinger.
7. Peele, S. (1985). *The meaning of addiction*. Lexington: Lexington Press.
8. Roques, B. (1998). *La dangerosité des drogues*. Paris : Odile Jacob, La Documentation Française.
9. Solal, J. F. (1997). Abstinence et substitution : histoire et actualité des traitements aux toxicomanes. *Pratiques Psychologiques*, 4, 13-24.

3) Numérotation incrémentale : ici, au fur et à mesure que les références apparaissent dans le texte, vous leur donnez un numéro croissant. Ensuite, la liste bibliographique est classée par numéro, ce qui signifie que les références bibliographiques ne sont pas listées dans l'ordre alphabétique d'auteurs. Ici aussi, l'ajout de références, après coup, décale complètement le

classement et est à la source d'erreurs (sans compter qu'une consultation libre de la liste de références est fortement gênée par l'absence d'ordre alphabétique des auteurs).

Exemple, méthode 3 :

Dans le texte…

Jusqu'à présent l'œuvre de Stanton Peele a été présentée dans la littérature française comme « éclectique et descriptive, empruntant au constructivisme social et recourant à des concepts comportementaux et cognitifs » (1). (…). Les débats des spécialistes à propos de la toxicomanie, aux États-Unis comme en Europe (2, 3, 4, 5) et les très récents rapports ministériels (6, 7), montrent que les ouvrages de Peele (8, 9) et l'ensemble de son œuvre, n'ont pas encore perdu une once d'actualité et qu'il peut être utile de les examiner à nouveau.[13]

Dans la liste bibliographique finale…

1. Pedinielli, J.-L., Rouan, G. & Bertagne, P. (1997). *Psychopathologie des addictions*. Paris : Presses Universitaires de France, Nodules.
2. MacCoun, R. & Reuter P. (1997). Interpreting dutch cannabis policy: Reasoning by analogy in the legalization debate. *Science*, 278, 47-52.
3. Henrion, R., et al. (1995). Rapport de la commission de réflexion sur la drogue et la toxicomanie. Paris : La Documentation Française.
4. Solal, J. F. (1997). Abstinence et substitution : histoire et actualité des traitements aux toxicomanes. *Pratiques Psychologiques*, 4, 13-24.
5. Lucas, B. (1996). Politique française en matière de drogue, la singularité du cas lyonnais. *Psychotropes*, 2, 75-95.

[13] Ibidem.

6. Roques, B. (1998). *La dangerosité des drogues*. Paris : Odile Jacob, La Documentation Française.
7. Maestracci, N. (1999). Lutte contre la drogue et la toxicomanie. Rapport remis au Premier ministre Lionel Jospin, janvier, confidentiel.
8. Peele, S. & Brodsky, A. (1975). *Love and addiction*. New York: Taplinger.
9. Peele, S. (1985). *The meaning of addiction*. Lexington: Lexington Press.

C'est évidemment la **méthode 1** que nous vous conseillons dans tous les cas. N'utilisez les autres méthodes que si la revue dans laquelle vous voulez publier (ou votre enseignant responsable) le demande.

Pour écrire les appels aux références bibliographiques dans le texte, selon la méthode 1, suivez les exemples ci-dessous (dans le dernier exemple, plusieurs références sont séparées par un point-virgule) :

- (Peele, 1985)
- (Peele & Brodsky, 1975)
- (Henrion, et al., 1995)
- (Browell, Marlatt, Lichtenstein & Wilson, 1986)
- (Peele, 1985 ; Peele & Brodsky, 1975 ; Henrion, et al., 1995 ; Browell, Marlatt, Lichtenstein & Wilson, 1986)

En ce qui concerne les normes de rédaction des références bibliographiques en liste, elles sont… trop nombreuses ! Nous vous conseillons de suivre la plus courante et

internationalement reconnue, le style APA (*American Psychological Association*) :[14]

Pour un article :

Murgatroyd, S., Rushton, C., Apter, M.J., & Ray, C. (1978). The development of the Telic Dominance Scale. *Journal of Personnality Assessment*, 42(5), 519-528.

Pour un ouvrage :

Apter, M. J., Kerr, J. H., & Cowles, M. P. (1988). *Progress in reversal theory*. Amsterdam: Elsevier.

Pour un chapitre dans un ouvrage :

Brown, R. I. F. (1988). Reversal theory and subjective experience in the explanation of addiction and relapse. In M. J. Apter, J. H. Kerr, & M. P. Cowles (Eds.), *Progress in reversal theory* (pp. 191-212). Amsterdam: Elsevier.

Remarques : soyez attentifs au jeu très subtil des espaces, points, virgules, parenthèses, italiques, etc., que vous devez respecter « à la lettre » ! Pour les chapitres dans un ouvrage, le (Eds.) qui indique les « éditeurs » de l'ouvrage, devient (Ed.) s'il n'y a qu'un seul éditeur.

2.9. La publication

En théorie, la publication des résultats de la recherche est la juste contrepartie de l'argent que la société donne au chercheur pour

[14] Évitez, sauf si votre enseignant responsable vous en donne la consigne, la norme franco-française, vilaine, vieillotte et étriquée ! Mais, de toute façon, pour publier vous devez vous plier aux normes de la revue et en tant qu'étudiant vous devez suivre les consignes de votre enseignant responsable.

qu'il fasse son travail de recherche. En réalité, l'étudiant, le thésard, non rémunérés, ou encore l'enseignant-chercheur surbooké par l'enseignement et sous-payé, n'ont pas à se sentir obligés de publier au profit d'une société qui les ignore ou les néglige.[15] En fait, la loi du *publish or perish* (publier ou crever !), toute égocentrique qu'elle soit, est malheureusement la principale motivation à la publication : publier pour grossir sa bibliographie personnelle, gonfler son C.V., son dossier de candidature aux concours et aux postes universitaires. Triste réalité, mais bon… c'est comme ça.

En ce qui concerne la publication dans une revue scientifique. Vous devez tout d'abord prendre connaissance des « instructions aux auteurs » et les lire attentivement afin de les suivre à la lettre. Ces instructions sont imprimées généralement à la fin des numéros des revues, mais vous pouvez aussi les trouver sur internet en recherchant le site officiel de la revue en question (utilisez l'infaillible moteur de recherche **Google**. De plus, **Google** comporte un outil de traduction qui peut vous donner une base de départ pour publier en anglais ; à partir de là, l'interprétation à l'aide de **ChatGPT** est pratiquement équivalente à avoir un anglophone à la maison. Selon votre niveau d'anglais cette interprétation avec l'IA est plus ou moins indispensable). Si vous suivez correctement ces instructions aux auteurs pour rédiger votre article, il ne devrait pas y avoir de problème, seule la qualité de contenu de l'article sera en cause.

Une chose très importante et qui doit faire l'objet de votre attention, c'est le titre de votre texte. Autant le titre d'un ouvrage

---

[15] Dans certains milieux privilégiés de la recherche (laboratoires, instituts, comme le CNRS ou l'INSERM) la situation est bien différente : avec un travail d'équipe, des collaborations internationales, des échanges incessants entre chercheurs, une concurrence entre équipes, un suivi au jour le jour des travaux, etc. Dans de tels contextes, la publication scientifique des travaux a encore du sens. Ce qui n'est pas le cas de la recherche universitaire, en France, notamment en psychologie…

peut être relativement « fantaisiste » et accrocheur, autant celui d'un article scientifique doit-il être le plus précis et explicite possible. En fait, le principe est qu'en lisant le titre on sache déjà de quoi parle l'article. Voici quelques exemples réels :

« Traitement de substitution par méthadone et buprénorphine pour la toxicomanie à l'héroïne. »

« Isolement sensoriel, étude psychopathologique et psychanalytique de la régression et du schéma corporel. »

"How part-time work intensity relates to drug use, problem behavior, time use, and satisfaction among high school seniors: Are these consequences or merely correlates?"

« L'alexithymie : évaluation, données quantitatives et cliniques. »

« Anxiété-trait, anxiété-état, changement réel, changement apparent. »

« Toxicomanie et délinquance, une question de style de vie ? »

"Metamotivational states of tennis players in a competitive situation: An exploratory study."

"Playing with emergency: A case study of reversal theory in artwork."

"The sensation of risk and motivational tendencies in sports: An empirical study."

"Factors affecting reversals: A laboratory study."

"Gambling and problem gambling in Washington State: A replication study."

"A replication study of obsessional followers and offenders with mental disorders."

« Étude longitudinale de la qualité de vie d'une cohorte de patients psychotiques. »

« Alcoolodépendance et séropositivité virale C : étude multicentrique descriptive de 684 malades. »

Vous remarquerez, en fin de liste, comment certains titres évoquent clairement le niveau de la recherche : étude exploratoire (*exploratory study*), étude de cas (*case study*), étude empirique (*empirical study*), étude de laboratoire (*laboratory study*), étude réplicative (*replication study*, qui refait une étude ou une expérience antérieure), étude longitudinale, étude multicentrique descriptive… Vous pouvez indiquer dans le titre d'un article s'il y a eu randomisation, groupe de contrôle, le nombre de sujets étudiés, etc.

3. FAIRE DES STATISTIQUES

3.1. L'étude fictive

Pour vous permettre de vous entraîner aux statistiques sous la direction du manuel, nous avons créé les données d'une petite étude fictive (données contenues dans le fichier Excel « Data.xlsx » et que vous pouvez télécharger sur le site de l'auteur-éditeur).[16]

Cette étude consiste en un plan de recherche combiné (donc, qui compare deux groupes, avec une répétition de la mesure, un test et un retest). Imaginez un échantillon de 40 patients dans une clinique. Vous randomisez (par tirage au sort) deux groupes égaux de patients (20 + 20). Le premier groupe, dit « Expérimental » suivra un traitement psychologique durant 8 semaines. Le second groupe, dit de « Contrôle », n'aura aucun traitement durant ces 8 semaines. Ainsi, la VI est le traitement, avec deux niveaux : traitement, pas de traitement.

La problématique de cette étude fictive est le devenir de la dépression et de l'anxiété sous l'effet d'un traitement psychologique. Les deux VD mesurées seront l'anxiété et la

[16] *https://bit.ly/3KPB3UR*

dépression. La HAD est utilisée (*Hospital Anxiety Depression scale*), qui comporte deux facteurs : anxiété et dépression.

Les travaux statistiques entrepris, après préparation du dossier Excel, seront :

- Présentation des variables descriptives.
- Contrôle de l'équivalence des groupes.
- Contrôle de l'effet des variables descriptives.
- Comparaison des groupes entre eux à S1 (première semaine) et à S8 (huitième semaine).
- Comparaison de chaque groupe avec lui-même entre S1 et S8 (test-retest).

Si vous connaissez bien Excel, vous pouvez sauter la partie « **Le dossier Excel** » du manuel et vous rendre directement aux études statistiques avec SPSS. Cependant, pour préparer les données de votre propre étude, il est nécessaire d'en passer par Excel. La raison en est qu'Excel est bien plus pratique pour saisir vos données que SPSS. SPSS est aussi un tableur et permet de saisir des données, mais son ergonomie reste toutefois bien limitée à côté d'Excel. Ce dernier permet de plus de traiter le dépouillement des instruments d'évaluation (échelles, inventaires, questionnaires), à l'aide de « formules » appropriées (ce que ne permet pas SPSS). Enfin, il est si facile d'ouvrir ensuite un dossier Excel dans SPSS et d'y choisir la page voulue, qu'on ne voit pas la raison de se priver du logiciel de Microsoft (ou de tout autre tableur équivalent).

3.2. Le dossier Excel

3.2.1. Repérage de certaines commandes d'Excel

Excel de Microsoft est un tableur.[17] Vous devez déjà en connaître les rudiments, car nous n'allons pas vous apprendre ici à vous en servir. Simplement, nous allons juste pointer quelques commandes de menu que vous serez amené à utiliser.

Pourquoi utiliser Excel ? Bien qu'il soit possible de saisir des données directement dans le logiciel de statistiques SPSS, en passant d'abord par Excel, on se simplifie grandement la tâche, ceci grâce aux fonctions et macro-commandes performantes de ce logiciel. Examinons les principales commandes à connaître et qui vous seront utiles. Veuillez ouvrir Excel à vide. Il vous affichera alors un dossier vierge (Fig. 3).

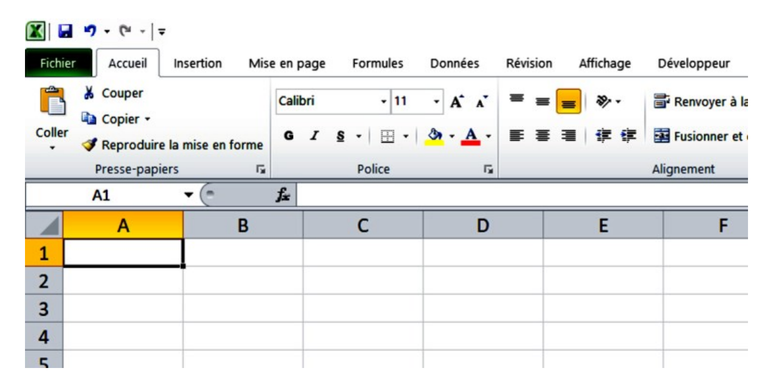


Figure 3 : Une feuille Excel vide de données.

Comment coller uniquement des valeurs numériques (et non les formules qui les déterminent) ?

Entrez dans la cellule A1 la formule suivante : **=22/7**, ce qui donne **3,14285714** comme résultat. Maintenant, vous allez copier-coller ce résultat, seulement sa valeur, pas la formule qui le donne, par exemple, dans la cellule **C1** (Fig. 4).

[17] Il s'agit de la version 2010, avec les « rubans » comme menus.

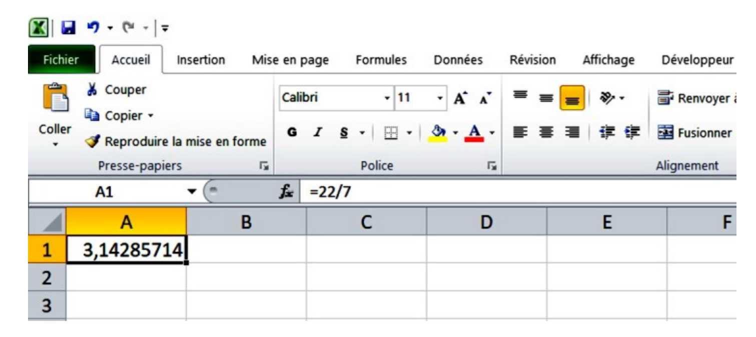


Figure 4 : Entrer une formule dans une cellule.

Vous allez utiliser le célèbre raccourci clavier **Ctrl+c** sur la cellule **A1** (ou bien vous faites un clic droit et vous sélectionnez l'option **Copier** dans le menu contextuel). Puis, vous ferez un clic droit sur la cellule de destination **C1**. Ce qui vous ouvrira un menu contextuel avec les options de collage (Fig. 5).

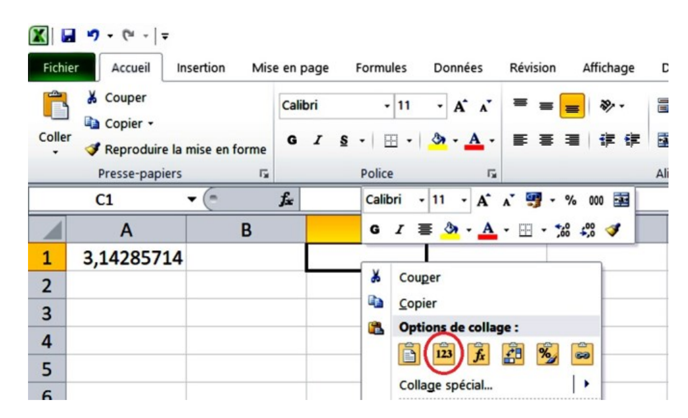


Figure 5 : Coller les valeurs uniquement.

Cliquez sur le bouton jaune **[123]** (cercle rouge), pour ne coller dans **C1** que le résultat de la formule (**3,14285714**) et non la formule elle-même (**=22/7**).

Comment copier vers le bas ?

Vous allez avoir besoin de copier-coller d'une façon rapide et pratique, une valeur ou une formule, d'une cellule en haut du tableur, vers les autres cellules au-dessous. Par exemple, entrez la valeur **123456** dans la cellule **K1**. Pour copier cette valeur de **K1** à

K6, sélectionnez de **K1** à **K6**. Puis, cliquez sur le ruban, onglet **Accueil**, sur la droite, dans le groupe **Édition**, sur le bouton **[Remplissage]**. Cliquez sur la première option du menu **En bas** (Fig. 6).

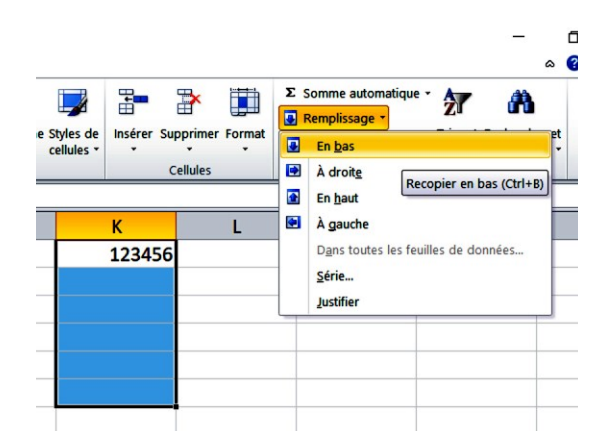


Figure 6 : Copier (remplir) vers le bas.

Cela marche aussi pour les formules, Excel prenant alors en compte les nouvelles données, de cellule en cellule. Dans **K1**, entrez la formule suivante : **=LIGNE(L1)**. Cette formule donne le numéro de la ligne, donc ici le résultat est **1**. Pour copier ce résultat vers les cellules au-dessous (de **K1** à **K6**, par exemple), au lieu de le faire manuelle, cellule par cellule, il vous suffit de sélectionner les cellules de **K1** à **K6**, puis de refaire le remplissage vers le bas, comme précédemment. Dans les cellules sélectionnées s'afficheront alors respectivement : **1, 2, 3, 4, 5, 6** (la formule prenant à chaque fois en compte la ligne au-dessous, de **L1** à **L6**, Fig. 7).

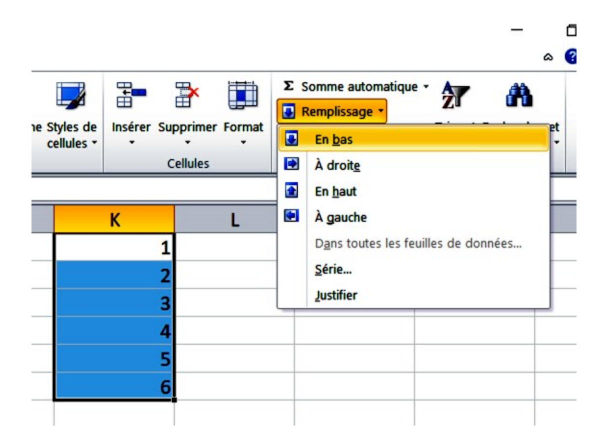


Figure 7 : Copier (remplir) vers le bas, résultat.

Vous pouvez réaliser les mêmes opérations, à partir du bouton **Remplissage**, vers le **Haut** en sélectionnant toujours plusieurs lignes, ou encore vers la **Droite** ou vers la **Gauche** en sélectionnant cette fois plusieurs colonnes.

Comment insérer ou supprimer des lignes et des colonnes ?

Sélectionnez la cellule autour de laquelle vous voulez ajouter une ligne (au-dessus) ou une colonne (à gauche). Puis, cliquez sur le ruban, onglet **Accueil**, dans le groupe **Cellules**, bouton **[Insérer]**. Dans le menu contextuel sélectionnez l'option **Insérer des lignes dans la feuille** ou **Insérer des colonnes dans la feuille** (Fig. 8-9).

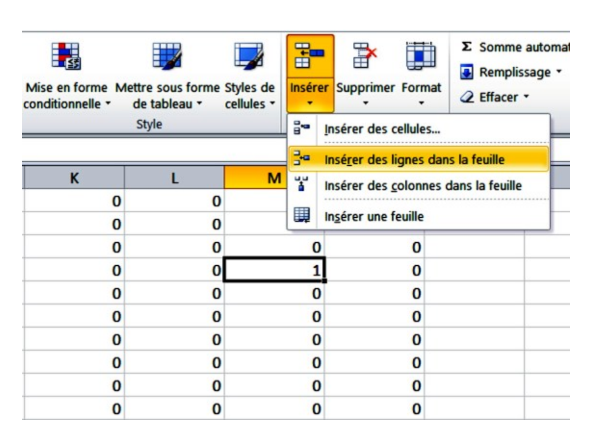


Figure 8 : Insérer des lignes et des colonnes.

Figure 9 : Insérer des lignes et des colonnes, résultat.

(Nota : dans la figure 9, la couleur bleue appliquée à la ligne et à la colonne insérées sert juste la démonstration, elle n'apparaîtra pas dans Excel). Pour supprimer lignes ou colonnes, utilisez le bouton **[Supprimer]** du ruban. Attention, seront supprimées les lignes ou les colonnes que vous avez sélectionnées.

Comment mettre et enlever le volet ?

Dans l'exemple présenté, vous avez entré les données des variables « Prénom », « Age », « Taille » et « Poids » (Fig. 10). Si vous avez beaucoup de variables (colonnes) et beaucoup de sujets de l'étude (les prénoms, un par ligne), il devient difficile de repérer les données (quelle variable pour quel sujet). Il est possible de figer, par exemple, la première ligne, celle qui contient le nom des variables, ce qui permettra de faire défiler vers le bas toutes les lignes des sujets, tout en maintenant la première ligne affichée.

Pareillement, en figeant la première colonne à gauche, celle qui contient les prénoms des sujets, vous pourrez faire défiler vers la droite les colonnes des différentes variables, tout en maintenant la première colonne affichée.

Presse-papiers | Police

H17 | f_x

	A	B	C	D
1	Prénom	Age	Taille	Poids
2	Bernard	33	1,72	85
3	Julien	45	1,7	78
4	Béatrice	28	1,56	75
5	Sandrine	26	1,63	63
6	Paola	32	1,68	62
7	Aziz	30	1,82	72
8	Paul	24	1,78	73
9	Fabienne	41	1,68	57
10				

Figure 10 : Un exemple fictif de quelques données.

Pour pouvoir ainsi fractionner le tableur, cliquez que le ruban **Affichage**, groupe **Fenêtre**, bouton **[Fractionner]**. Vous obtenez deux lignes poignets que vous pourrez déplacer à la souris, afin de fractionner sous la première ligne et à droite de la première colonne (Fig. 11-12).

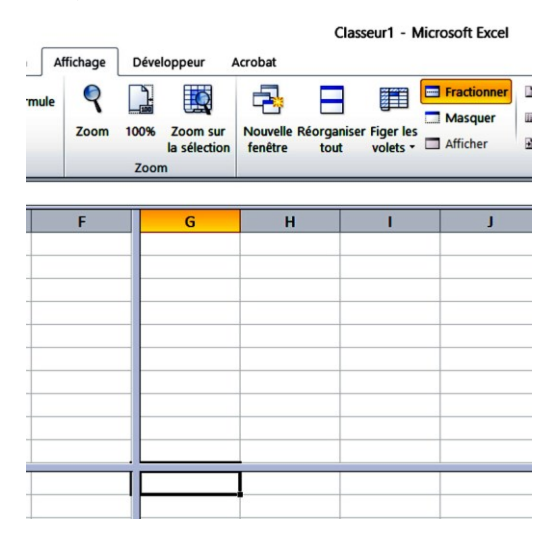


Figure 11 : Affichage des poignets de volets.

Presse-papiers | Police

J20 | fx

	A	B	C	D
1	**Prénom**	**Age**	**Taille**	**Poids**
1	**Prénom**	**Age**	**Taille**	**Poids**
2	Bernard	33	1,72	85
3	Julien	45	1,7	78
4	Béatrice	28	1,56	75
5	Sandrine	26	1,63	63
6	Paola	32	1,68	62
7	Aziz	30	1,82	72
8	Paul	24	1,78	73
9	Fabienne	41	1,68	57
10				

Figure 12 : Ajustement des poignets de volets.

Une fois les lignes poignets bien ajustées, vous allez les faire disparaître en cliquant sur le bouton **[Figer les volets]**, première option **Figer les volets** dans le menu contextuel (Fig. 13).

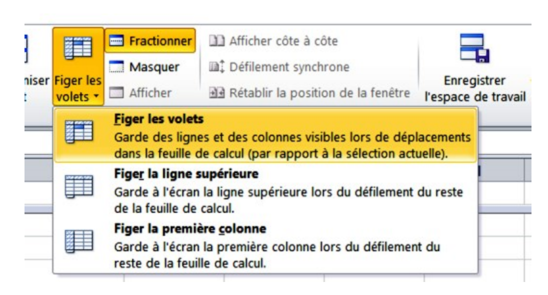


Figure 13 : Commande « Figer les volets ».

Les lignes poignets sont remplacées par de fines lignes noires (Fig. 14).

Affichages classeur | Affich

B23 | fx

	A	B	C	D
1	**Prénom**	**Age**	**Taille**	**Poids**
2	Bernard	33	1,72	85
3	Julien	45	1,7	78
4	Béatrice	28	1,56	75
5	Sandrine	26	1,63	63
6	Paola	32	1,68	62
7	Aziz	30	1,82	72
8	Paul	24	1,78	73
9	Fabienne	41	1,68	57
10				

Figure 14 : Affichage des volets figés.

Vous pouvez bien sûr faire machine arrière et déplacer les lignes de séparation ou supprimer les volets en cliquant sur les options correspondantes du ruban menu (bouton **[Figer les volets]**, option **Libérer les volets**, Fig. 15).

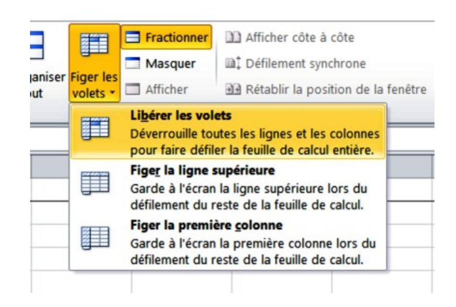


Figure 15 : Commande « Libérer les volets ».

Maintenant, vous êtes prêt pour ouvrir et examiner le dossier Excel « **Data.xlsx** » qui contient les données de l'étude fictive. Cette étude va vous permettre de bien comprendre tout le processus de traitement des données numériques dans le cadre de la recherche scientifique. « **Data.xlsx** » peut être aussi un modèle de dossier Excel que vous pourrez élaborer un jour avec vos propres données.

Cependant, si vous pensez avoir encore des choses à apprendre d'Excel, dans un premier temps, vous allez refermer « **Data.xlsx** » et partir d'un dossier Excel vide. Le but étant d'apprendre à construire un fichier de données, ce qui vous servira ensuite pour votre propre recherche, avec vos propres données.

3.2.2. Dossier Excel : la feuille principale

Avant tout chose, dans le dossier vierge créez une quatrième feuille (dans un dossier vierge, Excel vous propose d'emblée 3 feuilles), à l'aide du bouton **[Insérer une feuille de calcul]** (cercle rouge). Puis, en double-cliquant sur l'onglet de chaque feuille, vous pouvez les renommer. De gauche à droite les feuilles

s'appellent : « **Principale** », « **HAD** », « **Test-retest 1** » et « **Test-retest 2** » (Fig. 16).

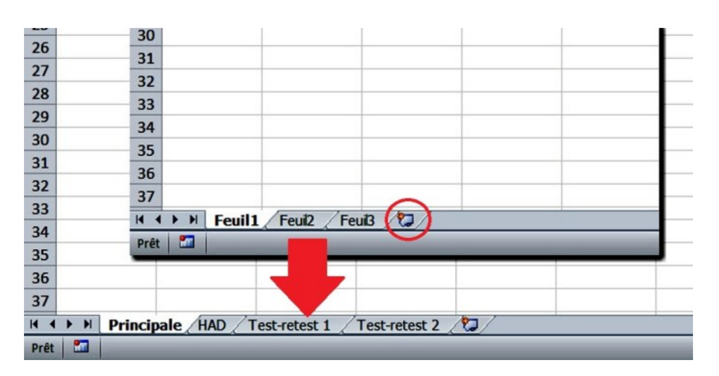


Figure 16 : Nommer les feuilles.

Positionnez-vous dans la feuille appelée « Principale ». En statistique, la présentation des tableaux de données se fait avec les variables en colonnes et les sujets en lignes. Première opération, créez la colonne « **numero** », elle permet d'identifier chacun de vos sujets par un numéro unique. Cela ne sert pas précisément pour la statistique, mais pour préparer le fichier Excel et pour éviter les confusions de données entre sujets (le numéro doit être reporté sur tous les documents, questionnaires, tests, qui concernent chaque sujet).

Colonne **A**, ligne **1**, entrez l'intitulé de colonne « **numero** » (important : pour les intitulés de colonne, ne donnait à SPSS que des noms de 8 caractères maximum, sans lettres accentuées, sans majuscules, sans point (.) ni tiret (-), sans caractères spéciaux, sans espace, juste lettres et chiffres). Dans la case en dessous entrez la formule « **=ligne(** » et cliquez sur la première cellule colonne B, « **B1** » s'inscrit entre les parenthèses. Appuyez sur la touche clavier **Entrée**. Le chiffre **1** s'inscrit dans la cellule **A2** et la parenthèse fermante est automatiquement ajoutée. Cliquez sur cette cellule : vous y voyez la formule « **=LIGNE(B1)** » (ovale rouge). Cette formule donne tout simplement le numéro de ligne de la cellule **B1**, Fig. 17).

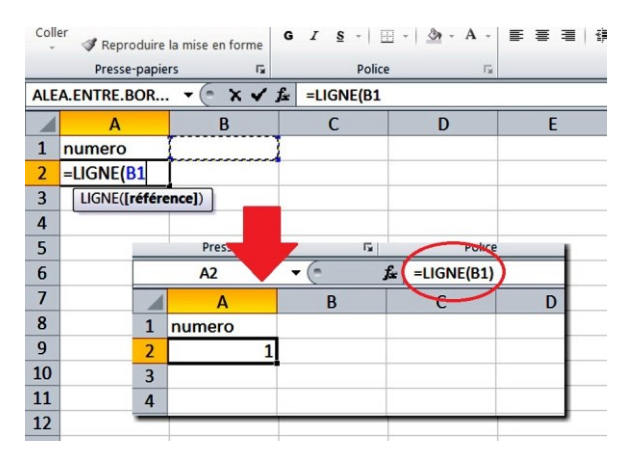


Figure 17 : Entrer une formule.

Pour copier vers le bas le numéro de ligne (donc le numéro des sujets de votre étude), sélectionnez à la souris la cellule **A2** et en maintenant enfoncé le bouton gauche de la souris, descendez vers le bas jusqu'à **A20**, par exemple. Puis, cliquez dans la barre d'outils sur le bouton copier vers le bas. La formule se reproduit de cellule en cellule, ce qui donne une numérotation croissante automatique (Fig. 18).

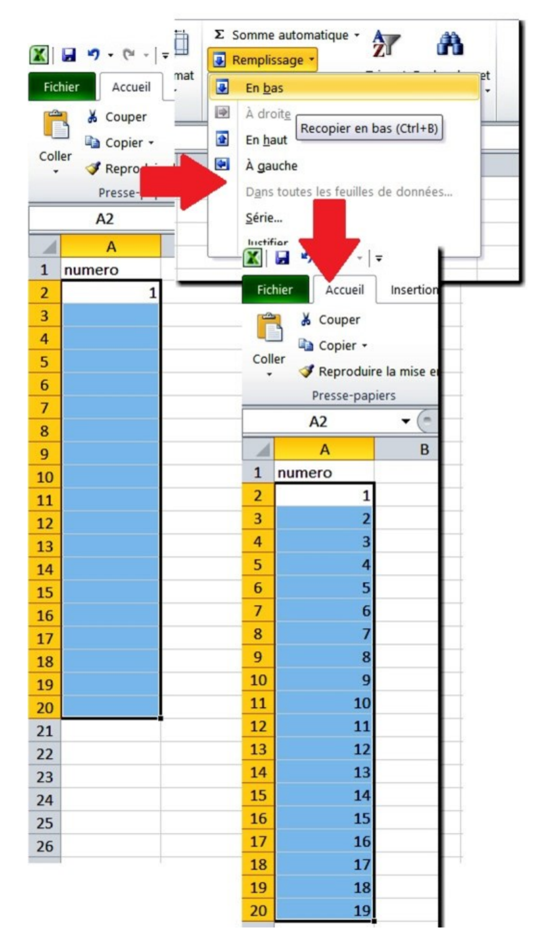


Figure 18 : Copier vers le bas la formule de numérotation.

Maintenant que vous avez compris, faites la même opération de **A20** jusqu'à **A41** afin d'avoir les numéros des 40 sujets de l'étude fictive (un sujet par ligne).

À présent, vous devez remplacer chaque formule « **=LIGNE(Bn)** » par sa valeur numérique, c'est-à-dire supprimer les formules « **=LIGNE(Bn)** » et ne garder que leurs résultats. Pour cela, sélectionnez les 40 cellules de **A2** à **A41** et faites un clic droit et sélectionnez l'option **Copier** (cercle rouge) dans le menu contextuel (Fig. 19, gauche).

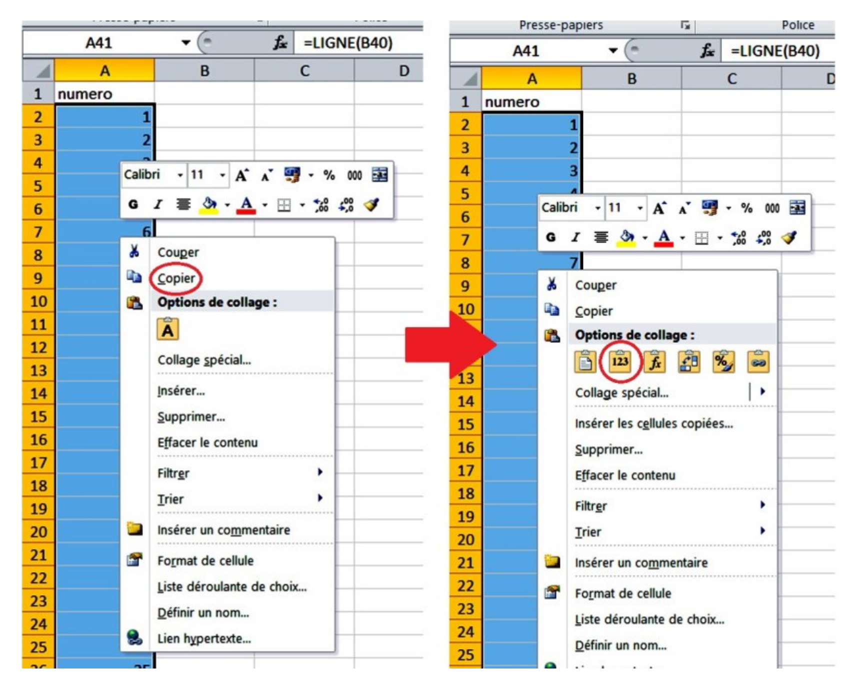


Figure 19 : Remplacer les formules par leurs valeurs.

Puis, vous refaites un clic droit et vous cliquez sur le bouton jaune **[123]** (cercle rouge). Ce bouton colle dans les cellules, non pas les formules mises en mémoire tampon, mais leurs valeurs, leurs résultats, de sorte que vous vous retrouvez avec simplement des nombres (ce qui fera plaisir à SPSS). Vérifiez-le, cliquez dans l'une des cellules de la colonne A, vous constaterez que la formule a disparu, vous ne voyez plus que sa valeur (Fig. 19, droite).

Créez à présent les intitulés des autres colonnes, pour les autres variables, en commençant par les variables descriptives et en terminant par les VD (Fig. 20).

- « groupe » (variable groupe Expérimental-Contrôle)
- « sexe » (variable sexe Homme-Femme)
- « age » (variable âge)
- « nivetude » (variable niveau d'études)

- « csp » (variable catégorie socioprofessionnelle)
- « hadanxi1 » (variable HAD anxiété test)
- « haddepr1 » (variable HAD dépression test)
- « hadanxi2 » (variable HAD anxiété retest)
- « haddepr2 » (variable HAD dépression retest)

	A	B	C	D	E	F	G	H	I	J
1	numero	groupe	sexe	age	nivetude	csp	hadanxi1	haddepr1	hadanxi2	haddepr2
2	1									
3	2									
4	3									

Figure 20 : L'intitulé des variables est entré en première ligne.

Vous remarquerez que nous respectons toujours la syntaxe SPSS pour le nom des variables. De plus, notez que nous vous conseillons d'utiliser la syntaxe suivante en ce qui concerne les VD : d'abord les initiales de l'instrument (had), puis les premières lettres du facteur concerné (anxi), jusqu'à la limite de 8 caractères. Comme nous l'avons fait, vous pouvez aussi intégrer un chiffre indiquant test ou retest (1 ou 2). Ce qui donne : had + anxi + 1 = « **hadanxi1** ». Cette façon de faire (voir la liste ci-dessus) permettra de ne pas vous tromper une fois sur SPSS (pour mémoire, prenez note sur papier ou dans un fichier texte de la signification de ce codage, par exemple : « hadanxi1 = HAD anxiété test », « hadanxi2 = HAD anxiété retest », etc.).

La saisie des données

Si vous travaillez avec l'étude fictive, vous pouvez « saisir » les données simplement en copiant-collant les données, à partir de « Data.xlsx », dans votre dossier vide (qui commence à se remplir). Dossier que pouvez d'ailleurs sauvegarder (nommez-le par exemple « MonData.xlsx »). Si vous voulez vous lancer avec vos propres données d'étude, vous pouvez commencer à les saisir

dans la feuille « Principale », au moins en ce qui concerne les variables descriptives (organismiques et démographiques).

Un conseil pratique important : n'attendez pas le dernier moment pour saisir vos données de recherche, à la fin, d'un coup. Pourquoi ? Parce qu'un tel travail devient vite fastidieux, pénible, ce qui peut occasionner des erreurs. De plus, à la fin de votre recherche, vous aurez besoin de toute votre énergie pour les analyses statistiques et rédiger le mémoire. Aussi, nous vous conseillons de saisir vos données au fur et à mesure des évaluations effectuées. Durant quelques semaines ou mois, vous entrerez de temps en temps les données d'un ou deux sujets et pour vous l'effort ainsi étalé sera imperceptible. Ne vous préoccupez pas de l'appartenance de groupe des sujets, entrez-les ligne par ligne comme ils viennent, nous verrons qu'Excel pourra très facilement faire un tri sur la colonne « groupe ».

L'utilisation du volet horizontal : il s'agit de bloquer à l'écran la première ligne, celle qui porte le nom des variables. Cette opération vous permettra, en saisissant vos données, de dépasser le bas d'écran tout en conservant la visibilité de l'intitulé des colonnes. Pour cela, dans le ruban menu **Affichage**, groupe **Fenêtre**, cliquez sur le bouton **[Fractionner]**.

Deux grosses lignes grises fractionnement alors l'écran horizontalement et verticalement. Débarrassez-vous de la ligne verticale (sauf si dans votre étude vous avez plus de 20 variables, donc plus de 20 colonnes à afficher, vous pouvez utilement fractionner verticalement en conservant l'affichage de « numéro » et « groupe », sur la gauche). Cliquez sur la ligne grise horizontale qui reste et placez-la juste sous la première ligne. Puis, dans le menu, cliquez sur le bouton **[Figer les volets]**, option **Figer les volets**. Vous obtiendrez une fine ligne noire séparant la première ligne des lignes de données au-dessous. Pour démonstration, descendez vers le bas, les dernières numéro

des sujets 37, 38, 39, 40. Vous constatez que la première ligne reste toujours bien affichée en haut (Fig. 21-22).

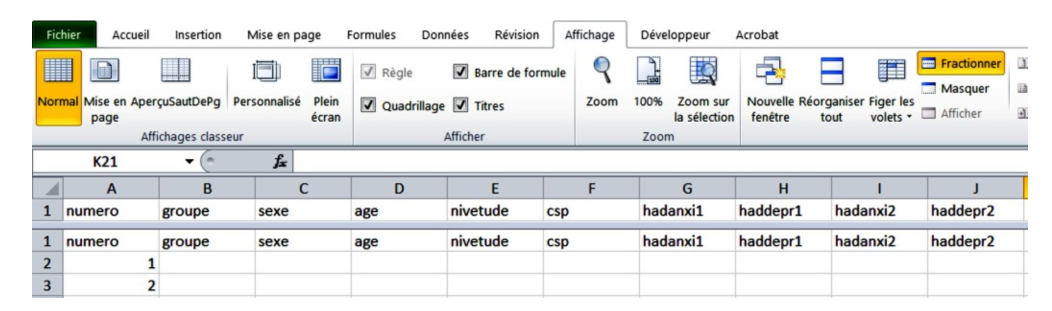


Figure 21 : Le volet est déplacé sous la première ligne.

Affichages classeur | Afficher | Zoom

L58

	A	B	C	D	E	F	G	H	I	J
1	numero	groupe	sexe	age	nivetude	csp	hadanxi1	haddepr1	hadanxi2	haddepr2
38	37									
39	38									
40	39									
41	40									

Figure 22 : Le volet est figé sous la première ligne.

Le codage des variables qualitatives-catégorielles

Bien que SPSS puisse réaliser certains calculs avec les valeurs alphabétiques des variables catégorielles (par exemple, avec les valeurs « Homme » et « Femme »), il est préférable de coder ces valeurs de façon numérique pour faciliter votre saisie. Pour cela, vous devez établir un codage, qui est libre et arbitraire, mais que vous devez consigner sur une feuille de papier ou dans un fichier texte, pour mémoire, afin de ne pas vous embrouiller une fois sur SPSS. Le codage que nous vous proposons est le suivant (nota : tout ceci est arbitraire, nous aurions pu prendre, par exemple, groupe « Expérimental » = 0 et groupe de « Contrôle » = 5).

- « groupe » (1 = Expérimental, 2 = Contrôle)
- « sexe » (1 = Homme, 2 = Femme)

- « nivetude » (1 = Primaire, 2 = Collège, 3 = Lycée, 4 = Université)
- « csp » (1 = Cadre, 2 = Employé, 3 = Ouvrier, 8 = Chômeur)

Nota : nous sommes dans un contexte fictif et les valeurs catégorielles attribuées sont fantaisistes (niveau d'études et CSP). Pour votre étude personnelle, référez-vous aux valeurs officielles.

Affichages classeur | Afficher | Zoom

A2 | f_x | 1

	A	B	C	D	E	F	G	H	I	J
1	numero	groupe	sexe	age	nivetude	csp	hadanxi1	haddepr1	hadanxi2	haddepr2
2	1	1	1	25	1	2	15	19	3	5
3	2	2	2	22	2	3	15	14	17	16
4	3	2	1	26	3	3	17	13	17	16
5	4	1	2	42	3	2	20	16	2	4
6	5	1	2	22	3	8	16	14	2	4
7	6	2	1	45	2	2	13	11	17	16
8	7	1	1	36	3	3	20	14	3	6
9	8	2	2	38	1	8	14	14	12	12
10	9	2	2	37	2	2	21	21	17	15
11	10	1	2	25	2	2	14	16	1	4
12	11	1	1	56	3	3	20	16	3	7
13	12	2	2	25	1	2	17	9	16	17
14	13	2	2	32	3	2	20	14	12	13
15	14	1	1	21	3	3	14	12	3	6
16	15	1	1	54	3	2	17	19	2	4
17	16	2	2	22	2	2	19	17	14	17
18	17	2	2	52	2	1	14	12	15	17
19	18	1	2	36	3	3	17	11	2	6
20	19	2	2	38	2	8	16	18	19	16
21	20	1	1	45	3	8	21	19	3	5
22	21	2	2	37	1	3	16	14	19	16
23	22	1	2	35	2	8	13	11	11	11
24	23	2	2	38	4	2	16	16	17	16
25	24	1	2	28	4	2	19	15	6	6
26	25	2	2	31	2	2	20	16	16	17
27	26	1	2	44	2	3	17	13	6	4
28	27	1	2	36	2	2	16	14	9	4
29	28	2	1	41	2	3	19	17	17	14
30	29	1	1	32	2	3	16	18	4	7
31	30	2	1	34	2	3	17	9	18	17
32	31	1	1	23	3	3	16	16	7	6
33	32	2	2	31	2	3	19	19	19	14
34	33	1	1	27	3	2	16	14	10	11
35	34	2	1	36	3	2	13	11	18	16
36	35	1	1	26	3	3	14	16	9	10
37	36	2	1	29	3	8	19	15	18	18
38	37	1	1	45	3	8	16	16	9	8
39	38	2	1	35	4	8	19	15	17	16
40	39	1	1	28	3	3	16	14	7	2
41	40	2	1	28	1	2	19	17	18	16
42										

Figure 23 : L'ensemble des données de Data.xlsx.

Les autres variables (âge et VD) sont quantitatives et continues, on entre simplement les âges et les scores obtenus aux

instruments (vous pourriez envisager de transformer l'âge en variable catégorielle ordinale, avec des tranches d'âge, mais à condition d'avoir un très grand nombre de sujets et qu'il y ait un intérêt à le faire, et pour l'étude et pour la problématique).

En résumé

Un tableau de données devrait comporter, de gauche à droite, les éléments suivants (voir illustration ci-après, correspondant à la feuille **« Principale »** du dossier d'étude fictive **Data.xlsx**, Fig. 23). Voici ce à quoi correspondent, exactement, les différentes variables de ce fichier.

- **numero :** c'est la variable d'identification avec un numéro. Une variable qui ne sert pas vraiment pour l'analyse statistique, mais pour le repérage et l'individualisation des sujets (contrôler la saisie, par exemple) et ainsi construire correctement le tableau des données. La colonne numéro permet aussi d'appliquer un tri sur d'autres variables (groupe, âge, sexe, valeurs croissantes d'une VD, etc.) puis de retrouver l'ordre initial en appliquant le tri sur la variable numéro). C'est aussi ce numéro que vous devez noter sur tous les documents papier de vos sujets (par exemple, un test), dans un souci de confidentialité.
- **groupe :** il s'agit de la variable indépendante (VI) principale de répartition des sujets en deux groupes, variable explicative principale, toujours qualitative, elle explique dans les hypothèses les valeurs obtenues aux VD (le cas échéant, vous pouvez aussi insérer sur la droite, d'autres colonnes de groupes, afin de tester d'autres modes de regroupement des sujets).
- **sexe, age, nivetud, csp :** il s'agit des variables descriptives, qui sont des variables indépendantes secondaires ou parasites. Certaines sont qualitatives, comme le niveau d'études, la CSP ; d'autres peuvent être quantitatives, comme l'âge, le nombre de sevrages antérieurs, le nombre d'enfants.

Ce sont ces variables dont il faut vérifier qu'elles sont équivalentes entre les groupes, afin de respecter le « toutes choses égales par ailleurs », c'est-à-dire que ces variables n'affectent pas les VD.

- **hadanxi1, haddepr1, hadanxi2, haddepr2 :** il s'agit des variables dépendantes (VD), qui sont généralement quantitatives. C'est la réponse des sujets aux instruments de mesure (questionnaires, tests, épreuves et tâches selon le dispositif expérimental mis en place). Parfois, ces variables peuvent être qualitatives-catégorielles, par exemple : vous transformez les scores continus à l'échelle de dépression en catégories ordinales (1 = pas de dépression, 2 = dépression légère, 3 = dépression, 4 = dépression grave). Cependant, vous devez savoir que ces catégories, si elles sont pertinentes dans une visée clinique (pour comparer un individu à une population de référence et poser ainsi un diagnostic), ne sont pas du tout pertinentes pour la recherche. En effet, vous allez ainsi perdre de l'information, car vous passez d'une fine variation continue à de grossières catégories qui regroupent, par exemple, tous les sujets qui ont eu un score entre 12 et 15.

3.2.3. Dossier Excel : la feuille HAD

Lorsque le score total d'un questionnaire se calcule simplement en additionnant le nombre de fois que le sujet a répondu « Oui », il reste facile de dépouiller un grand nombre de questionnaires. Par contre, si les sujets répondent à une échelle à 4 niveaux, si certains items sont inversés (il faut alors inverser le score à l'item : 0 devient 3, 1 devient 2, 2 devient 1 et 3 devient 0) et si, de plus, il vous faut calculer des sous-scores correspondants à autant de facteurs, le travail devient très vite fastidieux et source d'erreurs. Avec Excel vous pouvez grandement vous simplifier la tâche en lui demandant de faire les inversions de valeurs et les calculs pour vous. Ce que vous allez faire dans cette seconde feuille « **HAD** », du dossier Excel.

La HAD (*Hospital Anxiety Depression scale* – Échelle hospitalière d'anxiété et de dépression) comporte 14 items que le sujet est invité à coter entre 0 et 3 selon son « éprouvé au cours de la semaine qui vient de s'écouler ». Sept items concernent l'anxiété et 7 la dépression. Les items sont alternés : anxiété, dépression, anxiété, etc. Certains items sont inversés dans leur formulation : il faut inverser leur valeur comme cela est expliqué dans l'exemple au paragraphe précédent (à la HAD, l'item 4 de l'anxiété est inversé, ainsi que les items 1, 2, 3, 6 et 7 à la dépression). Pour construire une feuille Excel capable de tout calculer à votre place sans erreur (à condition vous-mêmes de ne pas faire d'erreur de saisie), voici la procédure.

1) Copiez-collez les deux colonnes « numero » et « groupe » de la feuille Principale, vers la feuille HAD. Ainsi, les deux premières colonnes de la feuille « **HAD** », en partant de la gauche, sont « **numero** » et « **groupe** ». Ensuite, tapez l'intitulé des deux colonnes suivantes, pour les résultats au test HAD : à l'anxiété (« **hadanxi** ») et à la dépression (« **haddepr** »). Laissez la colonne suivante vide, réduisez quelque peu sa largeur et donnez-lui une couleur de fond, cela afin de bien séparer visuellement les colonnes de résultats et celles des items qui vont venir.

En effet, à la suite vous entrez, toujours sur la première ligne, les intitulés des items : « **a1** » (= item 1 de l'anxiété), « **d1** » (= item 1 de la dépression), « **a2** » (= item 2 de l'anxiété), et ainsi de suite, jusqu'aux items **a7** et **d7**. Faites alors une nouvelle colonne vide et colorée comme séparation.

Les 6 colonnes suivantes concernent les formules d'inversion d'items. Pour l'anxiété, il faut inverser le score à l'item 4, vous intitulez la colonne « **inva4** ». Pour la dépression, il s'agit des items 1, 2, 3, 6 et 7, ce qui donne : « **invd1** », « **invd2** », « **invd3** », « **invd6** », « **invd7** ».

Maintenant, cliquez sur la cellule située juste sous l'intitulé de la colonne « **inva4** » et tapez (sans faire d'erreur) la formule conditionnelle suivante :

=SI(L2=3;0;SI(L2=2;1;SI(L2=1;2;3)))

Puis tapez la touche **Entrée**. Comment comprendre cette formule Excel ? D'abord L2 : si vous allez sur la gauche de la feuille, colonne L2, vous verrez qu'elle correspond à l'item « **a4** » (item 4 de l'anxiété, celui qu'il faut inverser). Pour lire la formule, vous dites en alternance « alors » et « sinon » à chaque point-virgule, ce qui donne :

si L2 = 3 alors L2 = 0 sinon si L2 = 2 alors L2 = 1 sinon si L2 = 1 alors L2 = 2 sinon L2 = 3, fastoche !

Une fois cette première formule écrite, sélectionnez la cellule qui la contient et en maintenant le bouton de la souris pressé, draguez vers la droite de la cellule U2 à Z2, puis, cliquez dans la barre d'outils les boutons **[Remplissage]**, puis **[À droite]**. Il ne vous reste plus qu'à modifier la cellule de référence pour chacune des formules. Pour « **invd1** », la cellule à tester est **G2**, pour « **invd2** » c'est **I2**, pour « **invd3** » c'est **K2**, pour « **invd6** » c'est **Q2** et pour « **invd7** » c'est **S2**.

Maintenant que les items inversés ont été pris en compte, il vous reste à établir les formules permettant d'additionner les points à l'anxiété et à la dépression. Pour cela, vous allez additionner les scores des items positifs et, lorsqu'un item est inversé, prendre son score inversé.

Pour l'anxiété, la formule à écrire dans la cellule **C2** est :

=SOMME(F2;H2;J2;U2;N2;P2;R2)

Pour la dépression, cellule **D2** est :

=SOMME(V2;W2;X2;M2;O2;Y2;Z2)

Les formules s'expliquent ainsi : il s'agit d'une formule Excel qui réalise la somme des valeurs des cellules placées entre les parenthèses, chaque identifiant de cellule étant suivie d'un point-virgule. Pour réaliser cela, en prenant l'exemple du score total à l'anxiété, il suffit d'entrer la formule de base dans la cellule **C2** : **=SOMME()**, de placer le curseur entre les parenthèses, puis de cliquer sur la première cellule à additionner, qui est **F2** (correspondant à l'item a1), **F2** s'inscrit automatiquement entre les parenthèses. Ensuite, maintenez enfoncée la touche du clavier Contrôle (**Ctrl**, en bas à gauche du clavier) et cliquez sur les cellules suivantes, une à une : **H2** pour l'item a2, **J2** pour l'item a3, attention **U2** pour l'item a4 sous sa forme inversée soit « **inva4** », **N2**, **P2** et **R2** pour les trois derniers items de l'anxiété, relâchez la touche **Ctrl** et tapez sur la touche **Entrée**. Même opération, en choisissant les cellules appropriées, pour le score total de dépression en cellule **D2**, « **haddepr** ».

À présent, la première ligne de formules est fonctionnelle, ce qui veut dire que si vous entriez les 14 cotations d'un sujet à la HAD, sur la première ligne sous les intitulés, entre a1 et d7, vous obtiendriez au niveau de « **hadanxi** » et « **haddepr** », les scores d'anxiété et de dépression. Pour préparer de la même façon les lignes pour les 39 autres sujets, il vous suffit de sélectionner la deuxième ligne de la colonne **F** à la colonne **Z**, puis tout en gardant pressé le bouton de la souris, de draguer vers le bas jusqu'à la ligne **81**, enfin de cliquer sur les boutons **[Remplissage]**, **[En bas]**.

Pourquoi jusqu'à la ligne **81** et non **41** ? Tout simplement parce que les sujets ont été testé deux fois. Aussi, vous allez placer à la suite : les 40 lignes pour le test des 40 sujets – de **2** à **41**, groupe Expérimental = 1 et groupe Contrôle = 2 ; et les 40 lignes pour le restest des mêmes 40 sujets – de **42** à **81**, groupe Expérimental = 11 et groupe Contrôle = 22. Vous avez à présent quatre groupes : deux groupes test (1 et 2) et deux groupes retest (11 et 22).

Dernière opération, tout comme pour la feuille « **Principale** », vous allez placer un volet horizontal juste sous la première ligne des intitulés (voir la page précédente), ainsi qu'un volet vertical. Cette fois, après avoir cliqué sur le bouton **[Fractionner]** et dragué avec la souris la barre grise horizontale juste sous la première ligne, vous allez draguer la ligne grise verticale juste à droite de la colonne au fond jaune **E**. Enfin, vous cliquez sur **[Figer les volets]**, option **Figer les volets** (Fig. 24-25).

Figure 24 : Les volets sont déplacés sous la première ligne et à droite de la colonne E.

Figure 25 : Les volets sont figés sous la première ligne et à droite de la colonne E.

La page de saisie et calcul des scores à la HAD est enfin prête, il ne vous reste plus qu'à entrer les cotations de vos sujets. C'était un exemple, un modèle, et cet exemple doit vous guider pour réaliser, pour vous, toutes les pages de dépouillement nécessaires selon les questionnaires, échelles, que vous utiliserez à l'avenir.

La fabrication d'une telle page de dépouillement prend un peu de temps, mais à l'usage, si vous devez dépouiller 50 protocoles, vous vous y retrouverez rapidement et surtout, quelle tranquillité car, sauf à vous tromper dans la saisie des cotations de vos sujets, si vos formules sont exactes, l'ordinateur ne fera aucune erreur de calcul et vous donnera immédiatement les résultats (pour en avoir la preuve : copiez-collez dans votre dossier Excel d'exercice, à partir du dossier « **Data.xlsx** » fourni avec le manuel, les données test-retest des sujets à la HAD, soit entre les colonnes **F** et **S** et les lignes **2** à **81** – feuille « **HAD** » naturellement. Dès le collage réalisé, vous verrez s'afficher les résultats dans les colonnes **C** et **D**, « **hadanxi** » et « **haddepr** », les scores qui seront utilisés dans SPSS).

3.2.4. Dossier Excel : les feuilles test-retest

Nous espérons que cet entraînement à la construction de feuilles de calcul Excel vous sera utile pour votre propre étude. À présent, nous ne pouvons plus avancer sans données, aussi, vous allez ouvrir dans l'application Excel le fichier « **Data.xlsx** ».[18]

Dans ce fichier « **Data.xlsx** », vous retrouvez les 4 feuilles : « **Principale** », « **HAD** », « **Test-retest-1** », « **Test-retest-2** ». Si vous les examinez, vous verrez qu'elles contiennent des données. Ces données fictives sont celles de l'étude fictive qui va nous servir d'exemple pour utiliser SPSS. Mais avant, vous devez comprendre comment sont organisées les feuilles.

[18] Pour rappel, vous pouvez obtenir le fichier de données sur le site de l'éditeur à l'adresse : *https://bit.ly/3yATjfv* (téléchargez le fichier « stats-pour-les-nuls-data.pdf » et renommez-le « stats-pour-les-nuls-data.zip ». Il ne vous reste plus qu'à ouvrir cette archive ZIP).

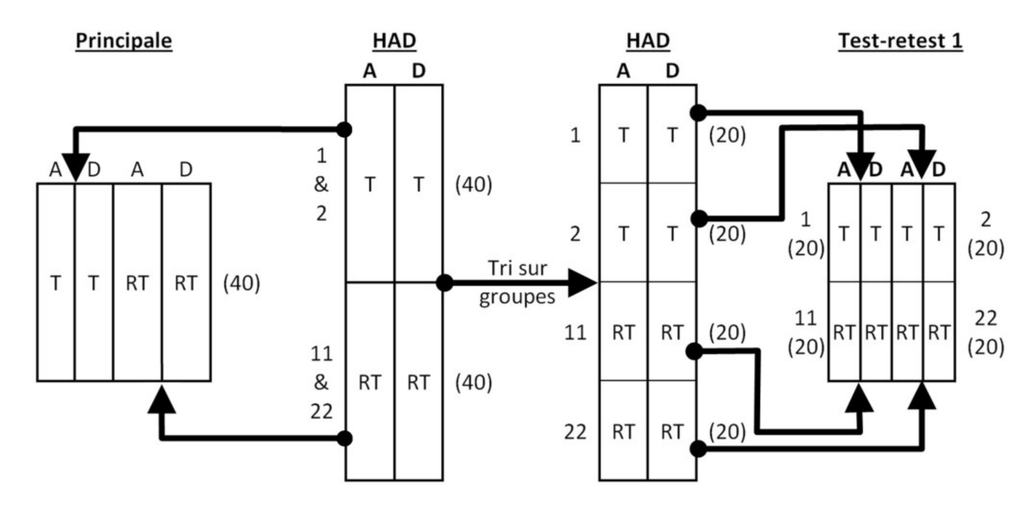


Figure 26 : Schéma des transferts de données de feuille en feuille.

Dans le schéma ci-dessus (Fig. 26) sont représentés l'organisation schématique des données et leur transfert d'une feuille à l'autre (flèches noires) par « Copier-coller » de leurs valeurs. Sur ce schéma, **A** = Anxiété et **D** = Dépression, **T** = Test et **RT** = Retest et entre parenthèses le nombre de sujets concernés : **20** (nombre de sujets dans un groupe) ou **40** (en prenant en compte les deux groupes de 20). Pour la feuille « **HAD** » groupe non trié, vous y reconnaissez les 40 tests (**T**), groupes 1 (Expérimental) et 2 (Contrôle) mélangés ; et en dessous les 40 retests (**RT**), groupes 11 (Expérimental) et 22 (Contrôle) mélangés.

À partir de la feuille « **HAD** », vous copiez-collez les « valeurs » (bouton jaune **[123]**) sur la feuille « **Principale** » (flèches noires partant vers la gauche) les données test (colonnes « **hadanxi1** » et « **haddepr1** ») et à côté, les données retest (colonnes « **hadanxi2** » et « **haddepr2** »). De cette façon, les résultats à la HAD, test et retest, sont bien enregistrés dans la feuille « **Principale** ».[19]

[19] C'est ce que vous pourriez faire dans votre dossier Excel d'exercice. Cela est évidemment, déjà fait, dans le dossier de données « **Data.xlxs** ».

Vous voyez sur le schéma que la feuille « **HAD** » subit une opération de tri sur les groupes. Ce tri est destiné à séparer les groupes « Expérimental » test et retest et « Contrôle » test et retest, afin de construire la feuille Test-retest. Pour réaliser le tri, vous devez tout d'abord sélectionner entièrement la feuille « **HAD** ». Pour cela il suffit de cliquer sur la case située en haut à gauche du tableur, à l'angle de la ligne 1 et de la colonne A (encadré rouge, Fig. 27).

	A	B	C	D	E	F	G
1	numero	groupe	hadanxi	haddepr		a1	d1
2	1	1	15	19		2	0
3	2	2	15	14		3	1
4	3	2	17	13		3	2
5	4	1	20	16		3	1

Figure 27 : Sélectionner entièrement la feuille « HAD ».

Toute la feuille doit alors passer en vidéo inverse. Ensuite, vous cliquez sur le ruban menu, onglet Données, groupe Trier et filtrer, bouton **[Trier]**.

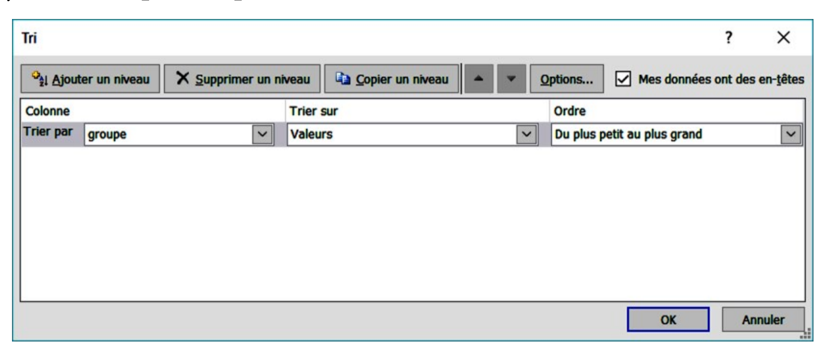


Figure 28 : Paramétrage de la fenêtre de « Tri ».

Dans la fenêtre « tri » (Fig. 28), vous cochez en haut à droite la case « Mes données ont des en-têtes » (ce qui veut dire que le tri se fera à partir des intitulés de variables sur la première ligne, sans que ces noms ne soient eux-mêmes pris en compte dans le tri).

Dans chaque liste déroulante, de gauche à droite, vous sélectionnez respectivement :

- **Colonne :** « groupe » ;
- **Trier sur :** « Valeurs » ;
- **Ordre :** « Du plus petit au plus grand ».

Enfin vous cliquez sur le bouton **[OK]**. À partir de ce tri, vous obtenez ce qui est représenté sur le schéma plus haut (Fig. 26) : les uns sous les autres, le test des 20 sujets du groupe « Expérimental » (1) et du groupe « Contrôle » (2), puis le retest des 20 sujets du groupe « Expérimental » (11) et du groupe « Contrôle » (22).

Nota : ce tri sur la feuille « **HAD** » est totalement réversible, grâce à la colonne « numéro ». En y appliquant un tri sur les valeurs, du plus petit au plus grand, vous allez retrouver l'état originel de la feuille.

Vous pouvez désormais réaliser la feuille « **Test-retest-1** » en copiant-collant les valeurs (concerne votre dossier Excel d'exercice) des données triées dans « **HAD** » (voir Fig. 26). Attention : pour les besoins de calcul de SPSS, concernant les deux groupes (« Expérimental » et « Contrôle ») les données test et retest sont rangées les unes sous les autres.

Explication : vous avez 20 lignes pour le test du groupe « Expérimental » (« grexpe » = 1) et dessous les 20 lignes pour le retest (« grexpe » = 11). Idem pour le groupe de « Contrôle », test pour « grcontr » = 2 et retest pour « grcontr » = 22.

Dans la feuille « **Test-retest-1** », vous avez ainsi deux colonnes groupe : « grexpe » (groupe expérimental, colonne « **A** ») et « grcontr » (groupe contrôle, colonne « **D** »). Chaque groupe comporte les résultats au questionnaire HAD (colonne « **B** », « anxiexpe » pour l'anxiété groupe expérimental et colonne « **C** », « deprexpe » pour la dépression groupe expérimental). Ces

résultats sont présentés ligne par ligne, de la ligne « **2** » à la ligne « **21** » pour le test et de la ligne « **22** » à la ligne « **41** » pour le retest. Et vous retrouvez la même présentation, mais cette fois dans les colonnes « **E** » et « **F** » pour le groupe de contrôle.

Pour réaliser la feuille « **Test-retest-2** », le schéma est le suivant (Fig. 29) :

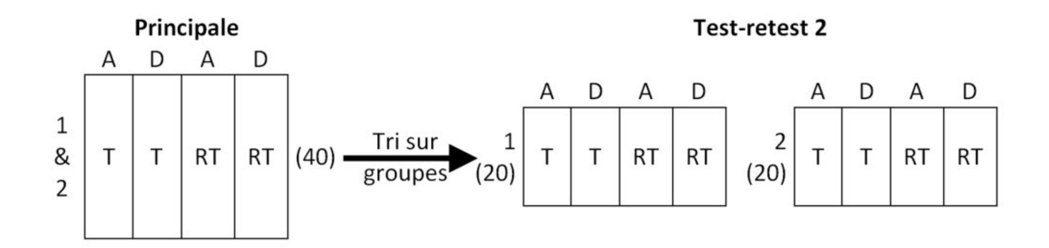


Figure 29 : Réalisation de la feuille « Test-restest 2 » à partir d'une copie de la feuille « Principale ».

La construction de la feuille « **Test-restest 2** » consiste à partir d'une copie de la feuille « **Principale** » (Fig. 30).

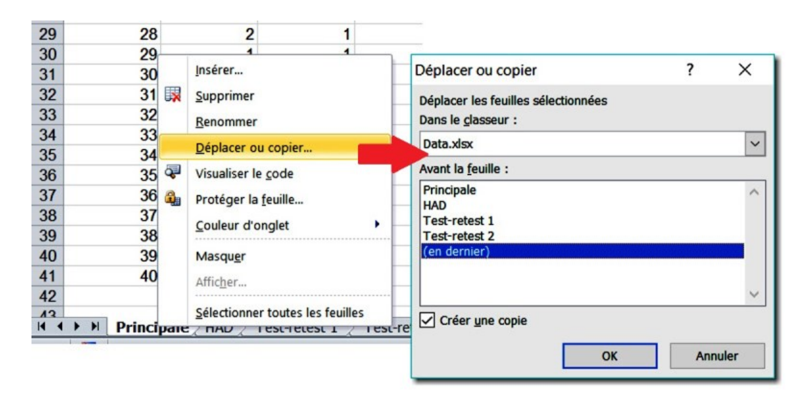


Figure 30 : Feuille « Principale » copier et renommée en « Test-restest 2 ».

Pour faire cette copie, vous commencez par un clic droit sur l'onglet de la feuille « **Principale** ». Dans le menu contextuel qui apparaît, cliquez sur l'option **Déplacer ou copier…**. Une petite fenêtre s'ouvre, dans laquelle vous allez sélectionner l'option **(en dernier)** (cela va placer la copie complètement à l'extrémité droite des onglets) et cliquer dans la case **Créer une copie**, enfin vous cliquez sur le bouton **[OK]**. Excel crée donc une copie qui s'appelle « **Principale (2)** », mais que vous allez tout de suite renommer en « **Test-restest 2** ».

Dans cette nouvelle feuille, vous supprimez les quatre colonnes des variables descriptives : « **sexe** », « **age** », « **nivetude** » et « **csp** » (pour ce faire, sélectionnez ces quatre colonnes – « **B** », « **C** », « **D** », « **E** » – puis vous cliquez ruban « **Accueil** », groupe « **Cellules** », bouton **[Supprimer]**, option **Supprimer des colonnes dans la feuille**.

Il ne reste que les variables « **numero** », « **groupe** » et les résultats à la HAD (« **hadanxi1** » et « **haddepr1** », pour l'anxiété et la dépression au test et « **hadanxi2** » et « **haddepr2** » au retest.

Appliquez ensuite un tri sur la variable « **Groupe** » (voir la méthode plus haut, Fig. 27-28), ce qui vous permettra de regrouper les données du groupe « Expérimental » (1) et celles du groupe de « Contrôle » (2), au-dessous. Pour la colonne « **A** », vous allez renommer l'intitulé « numero1 » et pour la colonne « **B** », ce sera « **grepex** ». Les intitulés des colonnes « **C** » à « **F** » restant inchangés.

Par contre, de « **G** » à « **L** », vous allez créer les intitulés suivants (qui vont reprendre, en les adaptant au groupe de contrôle, les intitulés du groupe expérimental) :

- Colonne « **G** » : « numero1 »
- Colonne « **H** » : « grcontr »
- Colonne « **I** » : « hadanxi3 »
- Colonne « **J** » : « haddepr3 »

- Colonne « **K** » : « hadanxi4 »
- Colonne « **L** » : « haddepr4 »

Puis vous allez déplacer les données du groupe de contrôle (2), de la ligne **22** à la ligne **41**, à la ligne **2**, colonnes « **G** » à « **L** ». Pour faire cela, vous sélectionnez toutes les données du groupe de contrôle

Retenez ainsi que, selon les calculs statistiques envisagés, SPSS a besoin qu'on lui place les données dans des colonnes différentes, ou bien dans les mêmes colonnes, les unes sous les autres. Il est temps, maintenant, de passer à SPSS.

En résumé, l'objectif de chaque feuille Excel est le suivant :

- **Feuille « Principale » :** Cette feuille regroupe l'ensemble des données, donc des variables : le numéro d'identification des sujets, leur appartenance aux groupes (VI), leurs caractéristiques démographiques (variables parasites) et leurs résultats aux tests (VD). Cette feuille servira à vérifier l'équivalence des groupes (par ailleurs) et à comparer les groupes sur les variables étudiées (VD).
- **Feuille « HAD » :** Il s'agit d'une feuille de dépouillement, qui permet de recueillir et calculer les scores aux instruments de mesure. Ses résultats sont transférés dans la feuille « **Principale** ».
- **Feuille « Test-retest 1 » :** Cette feuille présente les résultats aux tests (première évaluation) et va permettre de tester statistiquement l'évolution interne de chaque groupe entre test et retest.
- **Feuille « Test-retest 2 » :** Cette feuille reprend les résultats par colonne, afin de comparer (comparaison externe) les deux groupes entre test et retest.

3.3. Préparation de SPSS

Avant de vous lancer dans le traitement statistique, il vous faut comprendre et préparer SPSS.[20] SPSS est un logiciel de statistiques scientifiques performant. Avec ce présent manuel, nous n'utiliserons qu'une petite partie de ses possibilités. Vous devez avoir installé SPSS sur votre machine. Ouvrez SPSS pour l'examiner. Dès l'ouverture apparaît une fenêtre d'accueil (Fig. 31).

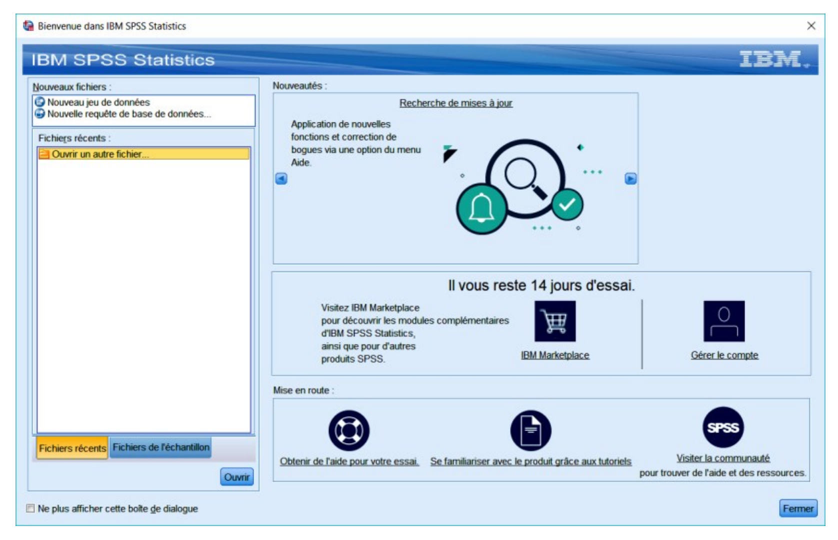


Figure 31 : Écran d'accueil de SPSS.

Vous pouvez, lors de cette première ouverture exploratoire et de réglage, fermer aussitôt cet écran (**nota :** si vous ne souhaitez plus le voir s'afficher au démarrage, pensez à cocher la case en bas à gauche « Ne plus afficher cette boite de dialogue ». Apparaît alors l'écran principal de SPSS (Fig. 32).

[20] Pour obtenir SPSS gratuitement, voir l'annexe en fin d'ouvrage.

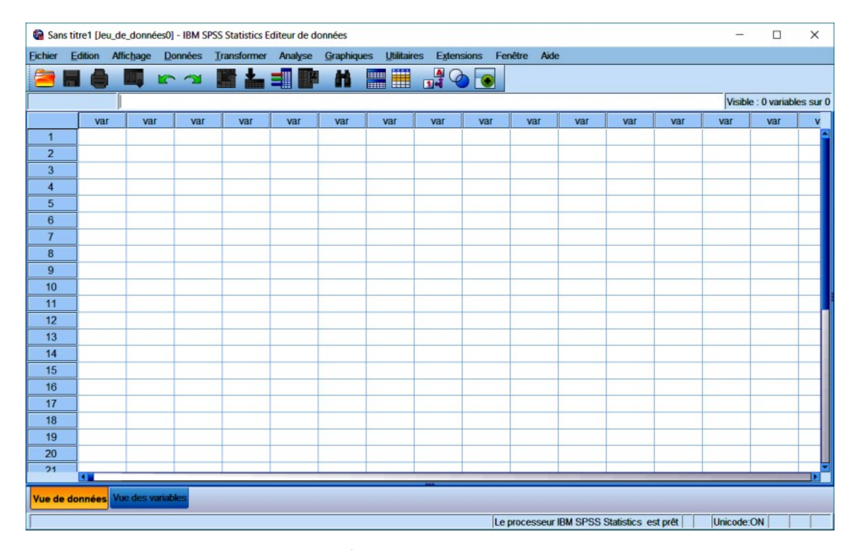


Figure 32 : Écran principal de SPSS.

À première vue, vous vous retrouvez comme dans Excel. SPSS est un tableur. Il y a un menu, une barre d'outils avec ses boutons, des lignes numérotées (on placera un sujet par ligne, comme dans Excel) et des colonnes pour les variables (d'ailleurs, l'en-tête des colonnes s'appelle déjà « **var** »).

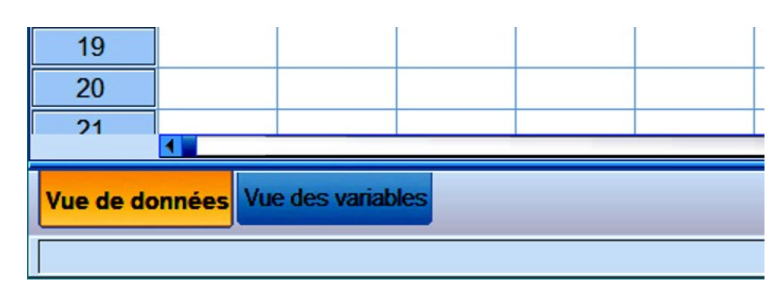


Figure 33 : Boutons des deux onglets de présentation des données et des variables.

Maintenant, si vous regardez en bas à gauche de la fenêtre SPSS, vous verrez deux onglets : **Vue de données** et **Vue des variables** (Fig. 33). Vous êtes actuellement dans l'onglet **Vue de données**, c'est la partie « tableur » de SPSS, qui permet de voir les

données. En cliquant sur l'onglet **Vue des variables**, vous pourrez voir et surtout modifier les caractéristiques de chaque variable. Ici, les variables sont en ligne et leurs caractéristiques (ce sont les variables des variables en quelque sorte) sont en colonnes (Fig. 34).

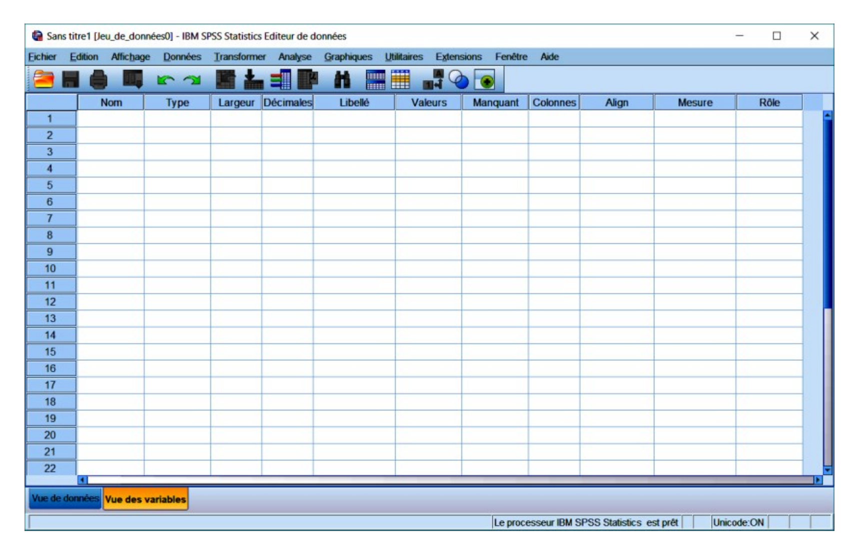


Figure 34 : Écran de présentation des variables.

Il y a 10 caractéristiques de variables, qui sont :

- **« Nom » :** le nom de la variable.
- **« Type » :** le type de la variable (de numérique à chaîne de caractères).
- **« Largeur » :** le nombre de caractères pour la donnée.
- **« Décimales » :** le nombre de décimales pour une variable numérique.
- **« Libellé » :** le titre de la variable (différent du nom, car plus explicite).
- **« Valeurs » :** les noms des niveaux de la variable qualitative (par exemple : 1 = Homme, 2 = Femme).

- **« Manquant » :** paramétrer les variables dont des données peuvent manquer.
- **« Colonnes » :** la largeur de la colonne du tableur.
- **« Align » :** l'alignement à l'affichage côté tableur.
- **« Mesure » :** sous-type de variable (échelle, nominal, ordinal).
- **« Rôle » :** non utilisé dans le cadre de ce présent manuel de statistiques.

La plupart des caractéristiques sont réglées automatiquement par SPSS lorsqu'il ouvre un fichier Excel (oui, c'est ce qui va bientôt arriver !). Nous verrons que seules, vont nous intéresser, les trois caractéristiques en souligné (« **Libellé** », « **Valeurs** » et « **Mesure** »). Avant de vous attaquer au paramétrage et réglage des options de SPSS, déplacez légèrement la fenêtre principale et vous verrez au-dessous une autre fenêtre (Fig. 35).

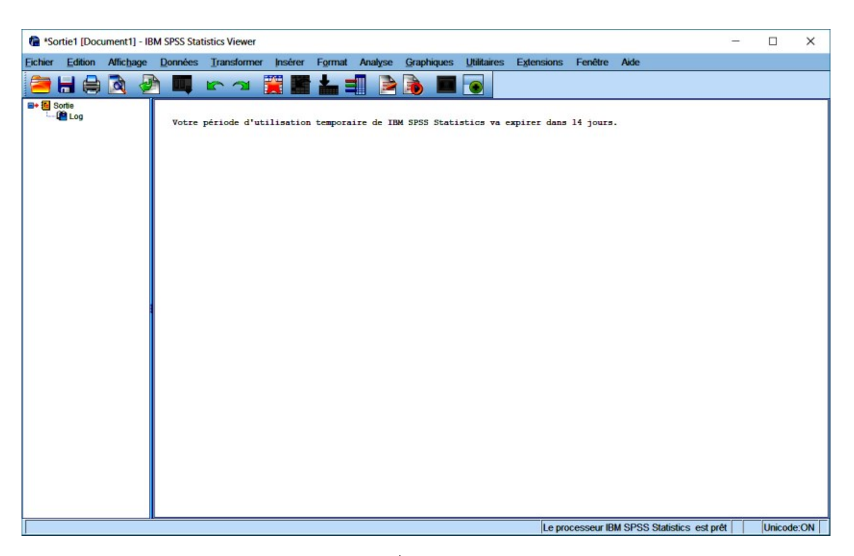


Figure 35 : Écran du Viewer.

Il s'agit de l'écran du **Viewer** (« Visualiseur »), qui s'est ouvert automatiquement au démarrage et qui servira à afficher et

enregistrer les résultats des calculs statistiques. À ce stade, vous pouvez fermer cet écran, pour le moment inutile.

Avant de commencer à travailler avec SPSS, vous devez régler certains paramètres. Pour cela vous allez cliquer dans le menu **Edition/Options** (tout en bas), ce qui vous donne la fenêtre ci-dessous (Fig. 36).

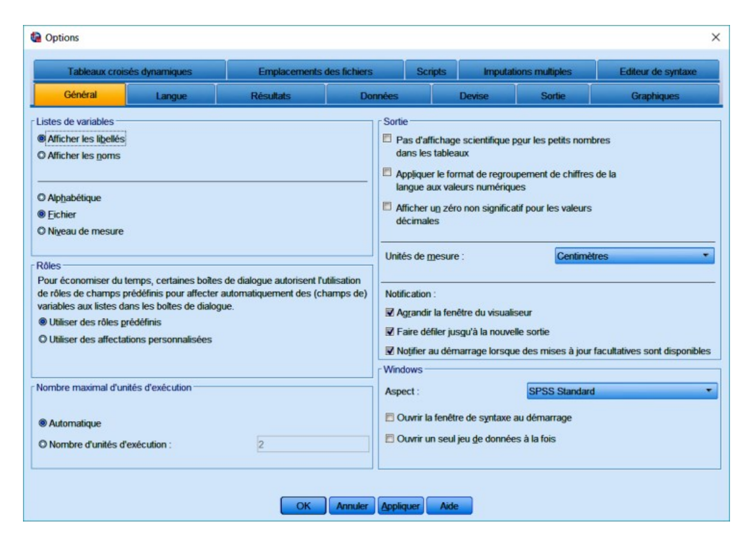


Figure 36 : Fenêtre des options.

Vous pouvez explorer un peu les différents onglets de cette fenêtre. Il s'agit ici de paramétrer le logiciel SPSS, pour lui donner certaines caractéristiques de fonctionnement ou d'apparence. En faisant des essais (allez-y, vous ne risquez rien d'essayer !), vous pourrez, par exemple, en allant sur l'onglet **Résultats**, déterminer les polices de caractère à employer pour afficher les résultats des calculs.

Ce qu'il est important de régler ici est le **type d'affichage des tableaux de résultats**. En effet, vous verrez bientôt que SPSS, dans sa magnifique toute-puissance, en deux-trois clics de souris, fait les calculs statistiques demandés et les affiche aussitôt. En science, il existe un format d'affichage pour les tableaux, dit « académique », qui correspond au standard

habituel de présentation (ouvrez une revue scientifique, vous verrez de quoi il s'agit). Pour programmer ce format, cliquez sur l'onglet **Tableaux croisés dynamiques** pour obtenir ceci (Fig. 37) :

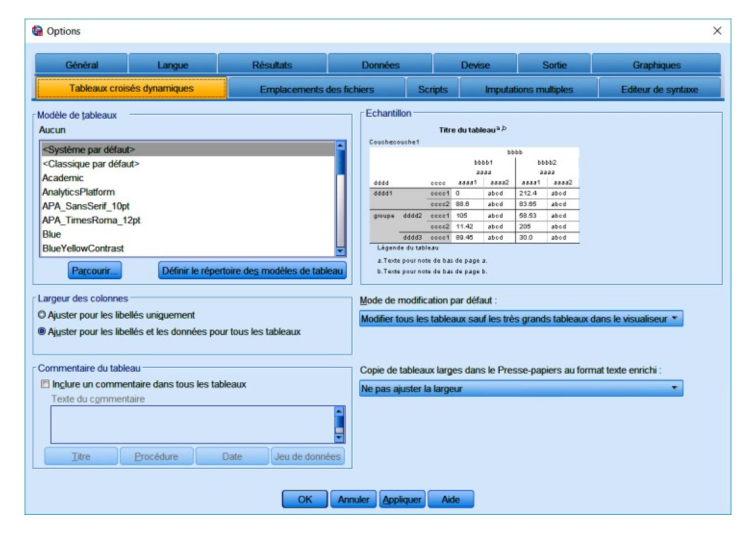


Figure 37 : Réglage de l'apparence des tableaux de résultat.

Dans la liste « **Modèle de tableaux** », choisissez le modèle « **Academic** » (**nota :** en fonction de vos besoins, vous pouvez, bien entendu, choisir le modèle qui vous convient). Cliquez sur le bouton **[OK]**, SPSS est prêt.

Nota : vous avez certainement remarqué que chaque fois que vous validez une modification d'option, SPSS ouvre une fenêtre du **Viewer** pour vous confirmer ce que vous venez de faire. Vous pouvez à chaque fois fermer cette fenêtre qui ne fait que vous donner un message sur l'opération accomplie et confirmer que vous ne voulez pas « Enregistrer le contenu du visualiseur de sortie » en cliquant sur **[Non]**. Pour rappel, c'est cette fenêtre « **Résultats** » (le **Viewer** en Anglais), qui s'ouvrira aussi pour vous donner les résultats des calculs statistiques de SPSS).

3.4. Création des fichiers SPSS

À partir du fichier de base « **Data.xlsx** » (d'Excel) vous allez créer trois fichiers au format **.sav** pour SPSS. Pour cela, vous ouvrez SPSS (si ce n'est déjà fait). Ensuite, vous cliquez dans le menu **Fichier/Ouvrir/Données**, ou plus simplement, vous cliquez sur le premier bouton à gauche de la barre d'outils **[Ouvrir un document de données]**. Vous obtenez la fenêtre « Ouvrir les données » ci-dessous (Fig. 38) :

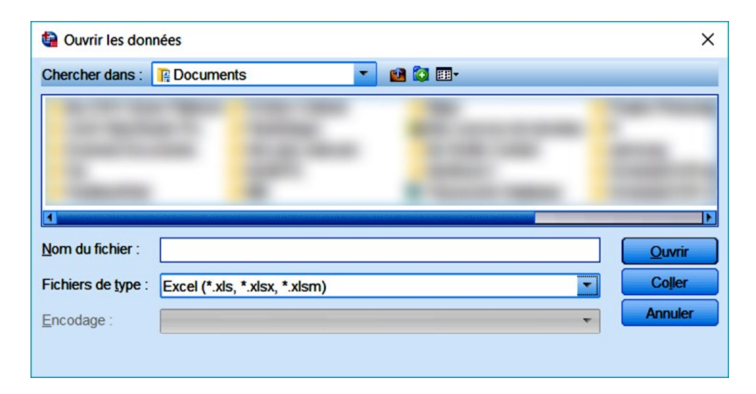


Figure 38 : Fenêtre d'ouverture des données.

Dans cette fenêtre, sélectionnez le type de fichier « Excel (*.xls, *.xlsx, *.xlsm,) » et, bien sûr, allez dans le répertoire où vous avez placé le fichier « Data.xlsx » et ouvrez-le.

Création de Principale.sav

Vous obtenez une petite fenêtre qui vous permet de choisir la feuille du dossier Excel (qui en contient plusieurs, comme vous le savez) que vous souhaitez utiliser dans SPSS (Fig. 39).

Sélectionnez la feuille « **Principale** » et cliquez **[OK]**. Dans l'onglet **Vue des variables**, du tableur, vous voyez le nom des 10 variables dans la colonne « **Nom** » : numero ; groupe ; sexe ; age ; nivetude ; csp ; hadanxi1 ; haddepr1 ; hadanxi2 ; haddepr2.

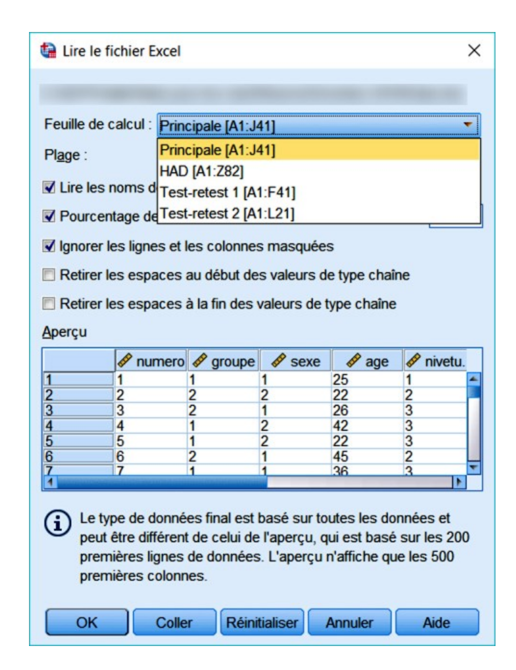


Figure 39 : Fenêtre de sélection de la feuille de données.

Tandis que dans l'onglet **Vue de données**, vous allez retrouver toutes les données, avec en têtes de colonnes les noms des variables, exactement comme dans Excel. Vous pouvez sauvegarder maintenant au format **.sav** (le format des fichiers SPSS) en cliquant sur le bouton **[Enregistrer ce document]** de la barre d'outils (le deuxième à gauche). Donnez un nom à ce fichier d'entraînement, par exemple « **Principale.sav** », en l'enregistrant sur votre disque dur.

Maintenant, cliquez à nouveau sur l'onglet **Vue des variables**. Les variables sont en lignes, leurs caractéristiques en colonnes. Seules trois caractéristiques sont intéressantes à modifier : « **Libellé** », « **Valeurs** » et « **Mesure** ». En renseignant « **Libellé** » et « **Valeurs** », au moment de l'affichage des résultats des calculs faits par SPSS, vous verrez alors, explicitement, les noms des variables et de leurs niveaux. Par exemple, fini « nivetude », niveaux 1, 2, 3 et 4, vous allez pouvoir paramétrer « **Niveau d'études** » : « **Primaire** », « **Collège** », « **Lycée** », « **Université** » et toutes ces informations seront clairement mentionnées dans les tableaux de résultats.

Pour « Libellé » (intitulé complet des variables), il suffit de cliquer à chaque fois sur la cellule de la colonne « Libellé » et de saisir le nom. Par exemple :

- Pour « numero » : **Numéro**
- Pour « groupe » : **Groupe**
- Pour « sexe » : **Sexe**
- Pour « age » : **Age**
- Pour « nivetude » : **Niveau d'études**
- Pour « csp » : **Catégorie socioprofessionnelle**
- Pour « hadanxi1 » : **HAD anxiété test S1**
- Pour « haddepr1 » : **HAD dépression test S1**
- Pour « hadanxi2 » : **HAD anxiété retest S8**[21]
- Pour « haddepr2 » : **HAD dépression retest S8**

Nota : pour écrire dans une cellule, il suffit de la sélectionner et de saisir votre texte. Mais pour modifier ce que vous avez écrit ou faire un copier-coller dans la cellule, vous devez d'abord faire un double-clic dessus pour la mettre en mode édition.

Notez les détails esthétiques : majuscules, lettres accentuées, pour que ce soit bien fait. À présent, passons aux noms des niveaux dans la colonne « **Valeurs** ». Ces noms des niveaux des variables n'intéressent, bien évidemment, que les variables catégorielles : le « Groupe », le « Sexe », le « Niveau d'études » et la « Catégorie socioprofessionnelle ». Pour faire cela, vous devez tenir compte du codage des niveaux choisi au moment de la saisie des données (si vous aviez choisi 1 = Homme et 2 = Femme, vous devez reprendre ce codage et non un autre ! Voir le chapitre « **Dossier Excel : la feuille Principale** », à « **Codage des**

[21] Pour rappel, l'étude fictive a fait un test à la première semaine (S1) et un retest à la huitième semaine (S8).

variables/Le codage des variables qualitatives-catégorielles »). Commençons par la variable « **Groupe** ». Dans la colonne « Valeurs », sur la ligne « Groupe », cliquez sur la cellule. Un bouton bleu apparaît avec **[...]**, sur lequel vous cliquez (Fig. 40).

	Nom	Type	Largeur	Décimales	Libellé	Valeurs	Manquant	Colonnes	Ali
1	numero	Numérique	2	0	Numéro	Aucun	Aucun	12	Droit
2	groupe	Numérique	1	0	Groupe	Aucun ...	Aucun	12	Droit
3	sexe	Numérique	1	0	Sexe	Aucun	Aucun	12	Droit
4	age	Numérique	2	0	Age	Aucun	Aucun	12	Droit

Figure 40 : Bouton pour attribuer les valeurs catégorielles.

Vous obtenez la fenêtre ci-dessous (Fig. 41) :

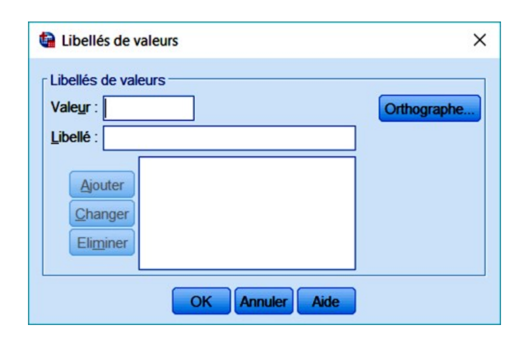


Figure 41 : Fenêtre d'attribution des valeurs catégorielles.

Dans la cellule « Valeur » entrez : **1**, et dans la cellule « Libellé » entrez : **Expérimental**, comme ci-dessous(Fig. 42) :

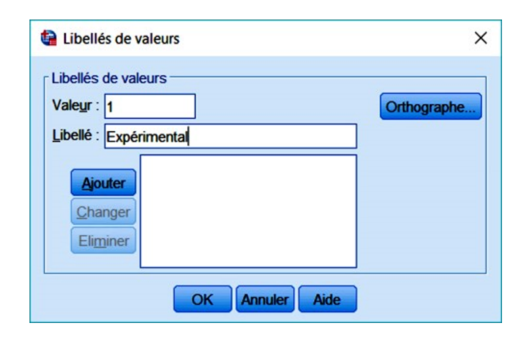


Figure 42 : Saisie de la correspondance valeur/libellé.

Puis cliquez sur le bouton **[Ajouter]**. Vous obtenez ce qui suit (Fig. 43) :

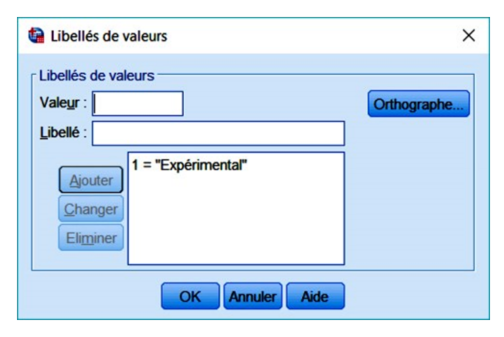


Figure 43 : Établissement de la correspondance valeur/libellé.

Faites de même pour « Valeur » : **2**, et « Libellé » : **Contrôle** et cliquez sur **[Ajouter]** (Fig. 44-45).

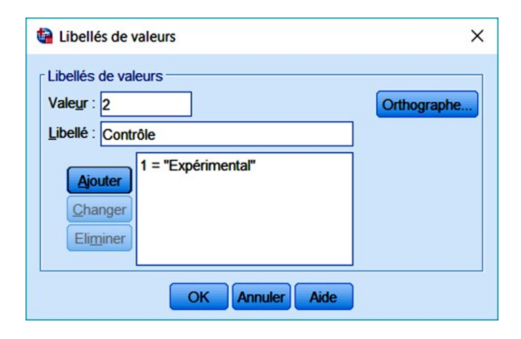


Figure 44 : Saisie de la correspondance valeur/libellé.

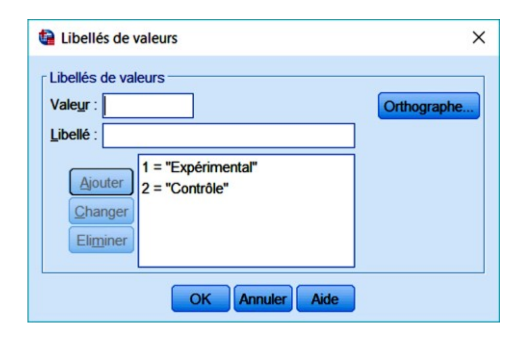


Figure 45 : Établissement de la correspondance valeur/libellé.

Il ne vous reste plus qu'à cliquer sur **[OK]**. Vous venez de nommer les deux niveaux de la variable « Groupe ». Il ne vous reste plus qu'à procéder de la même façon pour les autres variables qualitatives. Notez qu'en sélectionnant les correspondances « Valeur » et « Libellé » ainsi créées, vous pourrez ultérieurement les supprimer ou les modifier si besoin.

Rappel du codage des variables catégorielles

- **« groupe » :** 1 = Expérimental, 2 = Contrôle
- **« sexe » :** 1 = Homme, 2 = Femme
- **« nivetude » :** 1 = Primaire, 2 = Collège, 3 = Lycée, 4 = Université
- **« csp » :** 1 = Cadre, 2 = Employé, 3 = Ouvrier, 8 = Chômeur

La dernière opération va porter sur la caractéristique « Mesure » et ne concerne, ici aussi, que les variables catégorielles. Pour chacune d'entre elles (« **Groupe** », « **Sexe** », « **Niveau d'études** », « **Catégorie socioprofessionnelle** »), si SPSS ne l'a pas fait automatiquement pour vous, vous cliquez sur la partie droite de la cellule correspondante marquée « Echelle » et vous choisissez « Nominales », car il s'agit de variables nominales. « Echelle » concerne donc les variables numériques (par exemple, une valeur continue, comme l'âge ou le score à un instrument de mesure) et « Ordinales » concerne des variables ordinales (par exemple, dans « **Groupe** », 1 veut dire « Homme » et 2 veut dire « Femme »). Marquez toutes les autres variables avec « Echelle ». Au final, l'onglet **Vue des variables** doit donner ceci (Fig. 46)

Nota : « **Numéro** » est-il une variable d'échelle ou nominale ? Un numéro ne mesure rien, c'est juste une étiquette unique, un numéro par sujet étudié. Mais en même temps, comme aucun sujet n'a la même étiquette, numéro étant un identifiant unique, ce n'est pas non plus vraiment une variable nominale normale.

En fait, vous pouvez très bien laisser « **Numéro** » réglé sur « Echelle » (ou pas), car cette variable ne sera jamais utilisée pour les calculs statistiques dans SPSS.

*Principale.sav [Jeu_de_données1] - IBM SPSS Statistics Editeur de données

Fichier Edition Affichage Données Transformer Analyse Graphiques Utilitaires Extensions Fenêtre Aide

	Nom	Type	Largeur	Décimales	Libellé	Valeurs	Manquant	Colonnes	Align	Mesure	Rôle
1	numero	Numérique	2	0	Numéro	Aucun	Aucun	12	Droite	Echelle	Entrée
2	groupe	Numérique	1	0	Groupe	{1, Expérim...	Aucun	12	Droite	Nominales	Entrée
3	sexe	Numérique	1	0	Sexe	{1, Homme...	Aucun	12	Droite	Nominales	Entrée
4	age	Numérique	2	0	Age	Aucun	Aucun	12	Droite	Echelle	Entrée
5	nivetude	Numérique	1	0	Niveau d'études	{1, Primair...	Aucun	12	Droite	Nominales	Entrée
6	csp	Numérique	1	0	Catégorie soci...	{1, Cadre}...	Aucun	12	Droite	Nominales	Entrée
7	hadanxi1	Numérique	2	0	HAD anxiété te...	Aucun	Aucun	12	Droite	Echelle	Entrée
8	haddepr1	Numérique	2	0	HAD dépressi...	Aucun	Aucun	12	Droite	Echelle	Entrée
9	hadanxi2	Numérique	2	0	HAD anxiété r...	Aucun	Aucun	12	Droite	Echelle	Entrée
10	haddepr2	Numérique	2	0	HAD dépressi...	Aucun	Aucun	12	Droite	Echelle	Entrée

Figure 46 : Ensemble des caractéristiques des variables.

Nota : Concernant les caractéristiques de « Mesure » pour les variables, vous avez 3 possibilités (voir le chapitre « **1.4. Les types de variables** » (Fig. 47) :

- **Echelle :** Concerne les variables purement quantitatives, avec une valeur continue, comme l'âge (exemple : 1 an, 8 ans, 5 ans, 22 ans…).
- **Ordinales :** Concerne les variables quantitatives classées sous une forme qualitatives, catégorielle, comme les tranches d'âge (exemple : classe 1 entre 18-25 ans, classe 2 entre 26-35 ans, classe 3 entre 36-45 ans…). Ce peut être aussi une réponse à un questionnaire avec une **échelle de Lickert** (exemple : Pas du tout, Un peu, Moyennement, Beaucoup, Énormément).
- **Nominales :** Concerne des variables purement catégorielles, des sorte d'étiquettes (exemple : Homme ou Femme ; Étudiant, Chômeur, Travailleur…).

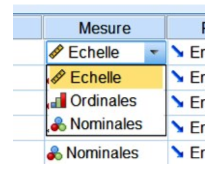


Figure 47 : Choix d'une caractéristique pour les variables.

Notez bien que dans la colonne « Type », toutes les variables sont numériques. Ce qui est vrai, car le fichier Excel au départ, ne contient que des nombres (par exemple, pour une variable catégorielle/nominale comme « groupe », les valeurs numériques sont 1 ou 2). Par contre, pour certaines variables, vous venez de les transformer de « numérique », en « catégoriel » et « nominal », en leur attribuant une valeur nominale (Expérimental, Contrôle), à la place de 1 et 2. Notez cependant que dès le dossier Excel vous auriez pu saisir « Expérimental » ou « Contrôle » au lieu de 1 ou 2, car SPSS est tout à fait capable de faire des calculs statistiques sur de telles étiquettes. Mais on ne procède pas ainsi, car il est plus facile et rapide, au moment de l'enregistrement des données dans Excel de taper, par exemple, « 1 » ou « 2 », plutôt que « Expérimental » ou « Contrôle ». Si vous avez une population de 50, 100 ou davantage de sujets, vous comprendrez vite pourquoi l'enregistrement dans Excel doit être de préférence numérique…

Vous avez terminé, sauvez une dernière fois votre fichier SPSS « Principale.sav » et passons au second fichier.

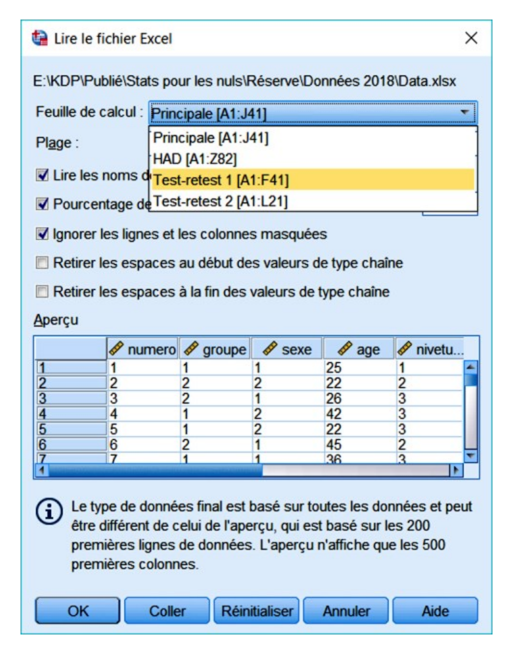


Figure 48 : Ouverture de la feuille « Test-restest 1 ».

Création de Test-retest-1.sav

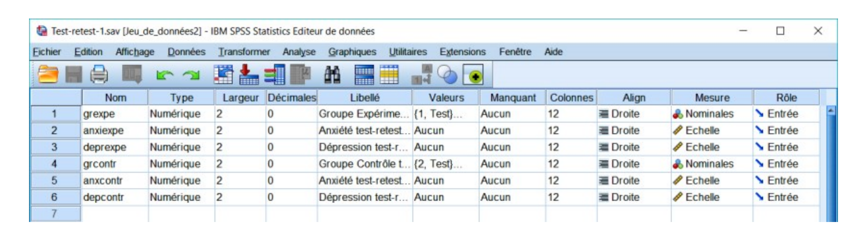
Recommencez à ouvrir le dossier « **Data.xlsx** ». Cette fois vous allez choisir la feuille Excel « **Test-retest 1** ». Vous allez au menu de SPSS **Fichier/Importer des données/Excel** et vous ouvrez « **Data.xlsx** ». Et dans « **Feuille de calcul** », vous sélectionnez « **Test-retest 1** » (Fig. 48).

Vous reconnaissez les mêmes dispositions que précédemment, il ne vous reste plus qu'à accomplir les mêmes opérations que pour « **Principal.sav** », dans l'onglet **Vue des variables**. Les libellés des variables sont ici :

- Pour « grexpe » : **Groupe Expérimental test-retest**
- Pour « anxiexpe » : **Anxiété test-retest Expérimental**
- Pour « deprexpe » : **Dépression test-retest Expérimental**
- Pour « grcontr » : **Groupe Contrôle test-retest**
- Pour « anxcontr » : **Anxiété test-retest Contrôle**
- Pour « depcontr » : **Dépression test-retest Contrôle**

Les noms des niveaux des variables qualitatives ne concernent ici que les groupes. Pour le groupe « **Expérimental** » (grexpe) on a : 1 = Test et 11 = Retest (colonne « Valeurs »). Pour le groupe « **Contrôle** » (grcontr) on a : 2 = Test et 22 = Retest.

Colonne « Mesure », donnez la valeur « Nominales » aux deux variables de groupe. Donnez la valeur « Echelle » aux autres variables. Sauvez ce nouveau fichier SPSS (par exemple, sous le nom très original de : « **Test-retest-1.sav** », Fig. 49).



	Nom	Type	Largeur	Décimales	Libellé	Valeurs	Manquant	Colonnes	Align	Mesure	Rôle
1	grexpe	Numérique	2	0	Groupe Expérime...	{1, Test}...	Aucun	12	Droite	Nominales	Entrée
2	anxiexpe	Numérique	2	0	Anxiété test-retest...	Aucun	Aucun	12	Droite	Echelle	Entrée
3	deprexpe	Numérique	2	0	Dépression test-r...	Aucun	Aucun	12	Droite	Echelle	Entrée
4	grcontr	Numérique	2	0	Groupe Contrôle t...	{2, Test}...	Aucun	12	Droite	Nominales	Entrée
5	anxcontr	Numérique	2	0	Anxiété test-retest...	Aucun	Aucun	12	Droite	Echelle	Entrée
6	depcontr	Numérique	2	0	Dépression test-r...	Aucun	Aucun	12	Droite	Echelle	Entrée
7											

Figure 49 : Ensemble des caractéristiques des variables.

Création de Test-retest-2.sav

Recommencez à ouvrir le dossier « **Data.xlsx** ». Cette fois vous allez choisir la feuille Excel « **Test-retest 2** » (Fig. 50).

Les étiquettes à attribuer aux variables sont les suivantes :

- Pour « numero1 » : **Numéro Expérimental**
- Pour « grexpe » : **Groupe Expérimental**
- Pour « hadanxi1 » : **Anxiété Test Expérimental**
- Pour « haddepr1 » : **Dépression Test Expérimental**
- Pour « hadanxi2 » : **Anxiété Retest Expérimental**
- Pour « haddepr2 » : **Dépression Retest Expérimental**
- Pour « numero2 » : **Numéro Contrôle**
- Pour « grcontr » : **Groupe Contrôle**
- Pour « hadanxi3 » : **Anxiété Test Contrôle**
- Pour « haddepr3 » : **Dépression Test Contrôle**
- Pour « hadanxi4 » : **Anxiété Retest Contrôle**
- Pour « haddepr4 » : **Dépression Retest Contrôle**

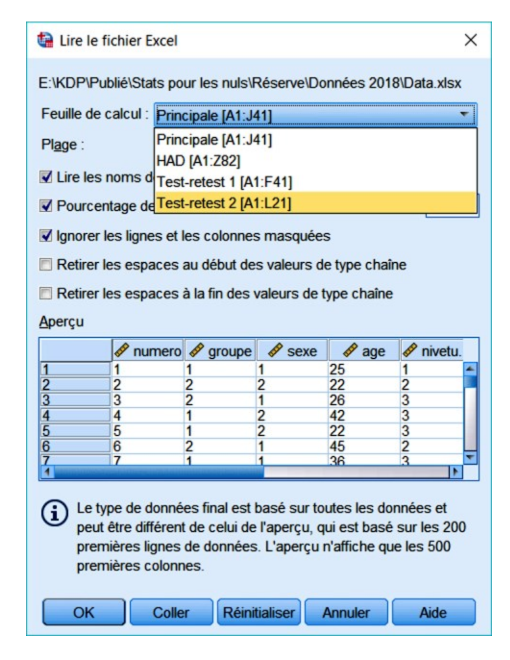
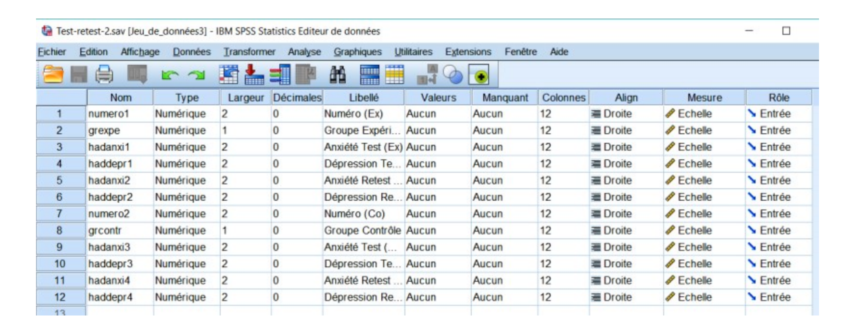


Figure 50 : Ouverture de la feuille « Test-restest 2 ».

Par ailleurs, vous n'avez aucun paramétrage de variable qualitative à faire, puisque de telles variables n'existent pas ici (« grexpe » et « grcontr » ne sont pas vraiment des « variables » car, justement, leur valeur ne varie pas – voir dans l'onglet **Vue de données**). Donc, dans la colonne « Mesure » attribuez la valeur « Echelle » à toutes les variables. Au final, l'onglet **Vue des variables** doit donner ceci (voir Fig. 51).



	Nom	Type	Largeur	Décimales	Libellé	Valeurs	Manquant	Colonnes	Align	Mesure	Rôle
1	numero1	Numérique	2	0	Numéro (Ex)	Aucun	Aucun	12	Droite	Echelle	Entrée
2	grexpe	Numérique	1	0	Groupe Expéri...	Aucun	Aucun	12	Droite	Echelle	Entrée
3	hadanxi1	Numérique	2	0	Anxiété Test (Ex)	Aucun	Aucun	12	Droite	Echelle	Entrée
4	haddepr1	Numérique	2	0	Dépression Te...	Aucun	Aucun	12	Droite	Echelle	Entrée
5	hadanxi2	Numérique	2	0	Anxiété Retest ...	Aucun	Aucun	12	Droite	Echelle	Entrée
6	haddepr2	Numérique	2	0	Dépression Re...	Aucun	Aucun	12	Droite	Echelle	Entrée
7	numero2	Numérique	2	0	Numéro (Co)	Aucun	Aucun	12	Droite	Echelle	Entrée
8	grcontr	Numérique	1	0	Groupe Contrôle	Aucun	Aucun	12	Droite	Echelle	Entrée
9	hadanxi3	Numérique	2	0	Anxiété Test (...	Aucun	Aucun	12	Droite	Echelle	Entrée
10	haddepr3	Numérique	2	0	Dépression Te...	Aucun	Aucun	12	Droite	Echelle	Entrée
11	hadanxi4	Numérique	2	0	Anxiété Retest ...	Aucun	Aucun	12	Droite	Echelle	Entrée
12	haddepr4	Numérique	2	0	Dépression Re...	Aucun	Aucun	12	Droite	Echelle	Entrée

Figure 51 : Caractéristiques des variables de « Test-restest 2 ».

Sauvez ce nouveau fichier SPSS (par exemple, sous le nom : « **Test-retest-2.sav** »). Les calculs vont pouvoir commencer, enfin !

Nota : vous vous étonnez peut-être : « Mais on ne fait rien avec la feuille « HAD » du dossier Excel ? ». Allons, il faut suivre ! La feuille HAD ne sert qu'à traiter, dans Excel, les réponses au test HAD anxiété-dépression auquel les sujets ont répondu. La feuille HAD n'est donc pas destinée à être traitée dans SPSS (dans le cadre de cette étude fictive).

En résumé, vous devez avoir désormais 3 fichiers SPSS (et le descriptif de leur prochaine utilisation) :

- **« Principale.sav » :** pour les statistiques descriptives, l'équivalence des groupes, l'effet des variables descriptives et les statistiques inférentielles.
- **« Test-retest-1.sav » :** pour les statistiques inférentielles.
- **« Test-retest-2.sav » :** pour les statistiques descriptives et les statistiques inférentielles.

3.5. Les statistiques descriptives

Nous l'avons vu au chapitre **Faire une recherche/Méthode**, à un moment donné il est nécessaire de décrire votre population d'étude, de comparer vos groupes, de montrer leur équivalence sur certaines caractéristiques et parfois leur différence significative sur une variable que vous contrôlez. Décrire de façon statistique c'est : indiquer des variables, présenter leurs effectifs (le nombre de sujets concernés par ces variables), donner les moyennes, les écarts types, les valeurs extrêmes minimum et maximum. Les statistiques descriptives concernent les variables organismiques (sexe, âge) et démographiques (niveau d'études, catégorie socioprofessionnelle, revenue, situation familiale, etc.). Le choix et le nombre de ces variables dépendent du type de recherche, des moyens disponibles et des objectifs de l'étude.

Nota : Pour tous les traitements statistiques vous allez utiliser avec SPSS les 3 fichiers **.sav** que vous venez de créer à partir du dossier Excel « **Data.xlsx** » (« **Principale.sav** », « **Test-retest-1.sav** » et « **Test-retest-2.sav** »).[22]

Groupe / sexe / âge

Ouvrir dans SPSS le fichier de données : « Principale.sav »

Décrire votre population selon les groupes qui la subdivisent,[23] le sexe et l'âge, est un classique et un incontournable. Sexe et âge sont deux variables « organismiques », il peut y en avoir d'autres, si nécessaire, comme l'appartenance ethnique, la couleur de la chevelure ou des yeux, le handicap, le niveau d'intelligence, etc.

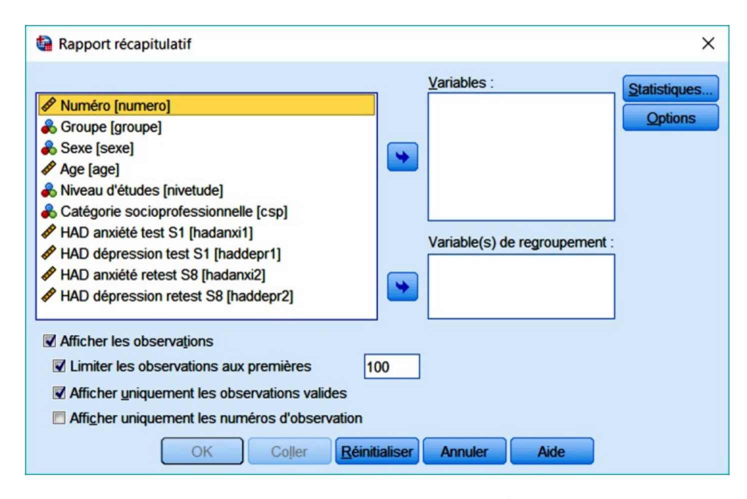


Figure 52 : Fenêtre « Rapport récapitulatif ».

[22] Si vous avez décidé de vous épargner l'apprentissage de la réalisation des fichiers SPSS **.sav** (ce n'est pas malin…), vous pouvez, bien entendu, utiliser directement les trois fichiers **.sav** fournis en téléchargement (*https://bit.ly/3KPB3UR*).

[23] Pour rappel, la variable « **Groupe** » est votre VI, c'est-à-dire la variable manipulée et explicative, par hypothèse, des résultats aux VD (les réponses aux questionnaires, par exemple).

Procédure SPSS :

Cliquez dans le menu : **Analyse/Rapports/Récapitulatif des observations**, vous obtenez la fenêtre suivante (Fig. 52).

Nous allons décrire cette première fenêtre de programmation de calculs statistiques. Nous ne le ferons plus par la suite, car toutes les fenêtres de SPSS, destinées à programmer des calculs, sont bâties, peu ou prou, sur le même modèle.

Dans la cellule de gauche, vous retrouvez la liste de toutes les variables du fichier SPSS ouvert. À droite, vous avez une ou plusieurs cellules dans lesquelles vous glissez les variables nécessaires aux calculs, ceci en sélectionnant à la souris la variable choisie à gauche et en cliquant sur le bouton flèche bleu approprié (le bouton au-dessous inverse naturellement le déplacement). Ensuite, vous avez divers boutons **[Statistiques....]**, **[Options]**, ou autres, selon les fenêtres. Boutons qui permettent de paramétrer les calculs à effectuer sur les variables choisies. Ce que vous devez faire à présent :

a. Glissez la variable « **Age** » dans la cellule « **Variables** » (ici l'on place la variable quantitative dont on veut présenter les valeurs).
b. Glissez les variables « **Groupe** » et « **Sexe** » dans la cellule « **Variable(s) de regroupement** » (ici l'on place les variables qualitatives de groupement sur lesquelles les différentes valeurs de la variable quantitatives sont réparties).
c. Décochez « **Afficher les observations** » (il est inutile d'afficher tous les sujets (les cas) en liste).
d. Bouton **[Options]** : dans la petite fenêtre, entrez le titre du tableau « **Tableau des caractéristiques Groupe/Sexe/Age** » dans la cellule « **Titre** », puis cliquez bouton **[Poursuivre]**.
e. Bouton **[Statistiques...]** : dans la petite fenêtre (dont vous pouvez régler les dimensions), de la cellule « **Statistiques** » à gauche, glissez dans l'ordre, dans la cellule « **Statistiques de**

cellules » à droite, les éléments suivants (dans l'ordre, c'est important) :

- Nombre d'observations (déjà présent à droite).
- Moyenne.
- Écart-type.
- Minimum.
- Maximum.

Puis, cliquez sur le bouton **[Poursuivre]**, vous venez de commander au menu de SPSS les calculs que vous souhaitez. Au final vous devez obtenir la fenêtre suivante (Fig. 53) :

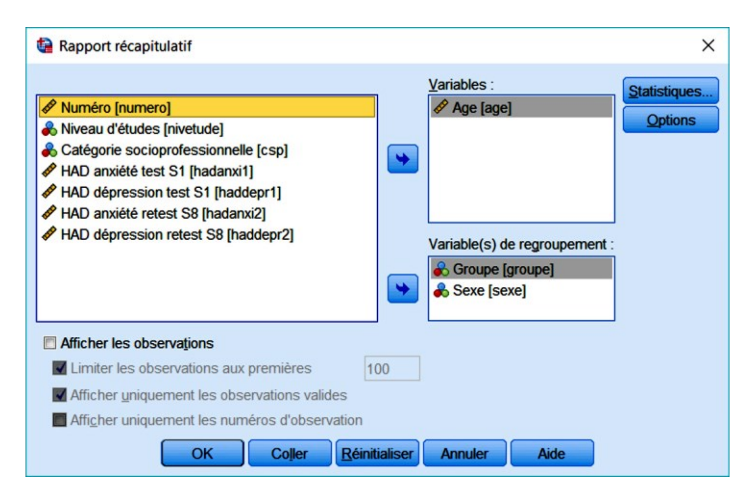


Figure 53 : Fenêtre « Rapport récapitulatif » paramétrée.

Il ne vous reste plus qu'à cliquer sur le bouton **[OK]**, SPSS exécutera les calculs demandés et affichera les résultats dans la fenêtre « Résultats » (cette fenêtre, le **Viewer**, ou « Visualiseur », est un programme annexe d'affichage, relié à SPSS – qui, lui, est un tableur et un calculateur de statistiques, qui s'ouvre dans une

seconde fenêtre et qui est destiné à afficher les résultats des calculs).

Le résultat est visible dans le tableau ci-dessous (Tab. 1, le premier tableau « **Récapitulatif de traitement des observations** » n'est qu'un tableau de contrôle qui vous permet de vérifier que toutes les données ont été prises en compte, qu'il n'y a donc pas d'anomalies dans ces données).

Tableau des caractéristiques Groupe/Sexe/Age

Age

Groupe	Sexe	N	Moyenne	Ecart type	Minimum	Maximum
Expérimental	Homme	12	34,83	12,231	21	56
	Femme	8	33,50	7,856	22	44
	Total	20	34,30	10,479	21	56
Contrôle	Homme	8	34,25	6,541	26	45
	Femme	12	33,58	8,415	22	52
	Total	20	33,85	7,541	22	52
Total	Homme	20	34,60	10,122	21	56
	Femme	20	33,55	7,983	22	52
	Total	40	34,07	9,014	21	56

Tableau 1 : Votre premier résultat avec SPSS.

Les possibilités du Viewer

Le (IBM SPSS Statistics) **Viewer** est un programme d'affichage complet, qui vous permet d'afficher les calculs issus de SPSS, d'ouvrir des fichiers de résultats au format **.spv** déjà créés et de sauvegarder les résultats sous ce format **.spv**. Vous pouvez, d'ailleurs, tout de suite, sauver le présent résultat descriptif en cliquant sur le bouton **[Enregistrer ce document]** de la barre d'outils (second bouton à gauche), donnez-lui, par exemple, le nom « **Statistiques-descriptives.spv** ».

Lorsque vous avez démarré SPSS avec un fichier de données, vous pouvez ouvrir un fichier **.spv** déjà existant, de sorte que les

résultats des prochains calculs se placeront à la suite dans ce même fichier **.spv**.

Le **Viewer** est séparé en deux volets (dont vous pouvez ajuster la taille) : le volet de droite où s'affichent les résultats et le volet de gauche qui présente l'arborescence (le menu) des différents résultats et qui vous permet d'y accéder facilement, comme dans la table des matières d'un livre.

Vous devez savoir que tout ce que vous voyez affiché dans ces deux volets peut très facilement être modifié ou supprimé. Il suffit de cliquer une fois sur l'élément (volet gauche), ensuite en appuyant sur la touche **Suppr** de votre clavier, vous supprimez cet élément. En cliquant une seconde fois sur l'élément ainsi sélectionné, il se met en position d'édition, ce qui signifie que vous pouvez le renommer. Cette possibilité de renommer, de modifier le texte, existe aussi dans les tableaux de résultats : vous double-cliquez sur le tableau, un encadrement apparaît, vous double-cliquez alors sur l'élément concerné, par exemple le titre du tableau et vous le modifiez à votre guise. Cliquez à l'extérieur du tableau et le mode édition disparaît.

Dans le fichier **Statistiques-descriptives.spv** qui vient d'être créé, vous pouvez par exemple supprimer le tableau « **Récapitulatif de traitement des observations** » : cliquez dessus (menu gauche), touche supprimer et c'est fait. Modifiez le titre « **Résumer** » (en gras) en double-cliquant dessus et saisissez le nouveau titre : « **Statistiques descriptives** ». Vous pouvez faire la même chose dans la partie gauche, afin d'avoir une table des matières explicite, dans laquelle il est facile de naviguer. Dans cette table des matières, vous pouvez très facilement, avec la souris, déplacer des éléments verticalement ou, pour certains, horizontalement (changement de niveau dans l'arborescence) en draguant avec la souris. Vous pouvez aussi ouvrir ou fermer cette arborescence, comme dans l'explorateur Windows, en cliquant sur les petits carrés + ou –.

Enfin, vous pouvez ajouter tout ce que vous voulez dans votre page de résultats, en allant au menu **Insérer**. Par exemple,

ajouter un titre, un en-tête, du texte, des images… Tout cela peut bien sûr être imprimé. Une particularité du **Viewer** : si vous ne sélectionnez aucun élément, tout sera imprimé ; par contre, si vous sélectionnez un ou plusieurs éléments, seuls ces éléments sélectionnés seront imprimés. Donc, attention, si vous ne voyez, par exemple, qu'un titre s'imprimer, c'est tout simplement parce que ce titre est sélectionné et que tout le reste est ignoré. Pour désélectionner, cliquez simplement à l'extérieur de tout élément. Voici ce que peut donner la fenêtre du **Viewer** après un petit nettoyage (Fig. 54).

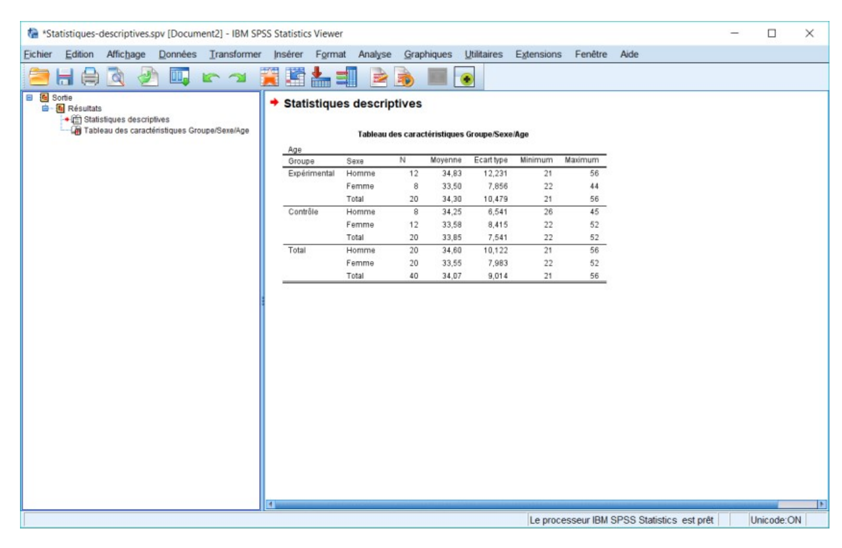


Tableau des caractéristiques Groupe/Sexe/Age

Age

Groupe	Sexe	N	Moyenne	Ecart type	Minimum	Maximum
Expérimental	Homme	12	34,83	12,231	21	56
	Femme	8	33,50	7,856	22	44
	Total	20	34,30	10,479	21	56
Contrôle	Homme	8	34,25	6,541	26	45
	Femme	12	33,58	8,415	22	52
	Total	20	33,85	7,541	22	52
Total	Homme	20	34,60	10,122	21	56
	Femme	20	33,55	7,983	22	52
	Total	40	34,07	9,014	21	56

Figure 54 : Présentation des résultats dans le Viewer.

Poursuite des statistiques descriptives

Remarque : pour le premier calcul, vous êtes parti du tableur SPSS. Mais à présent, vous pouvez fort bien réaliser de nouveaux calculs indifféremment, soit du tableur (onglet « **Vue de données** » ou « **Vue des variables** »), soit du **Viewer**. Cette possibilité est, aussi bien, valable pour les boutons de la barre d'outils (toutefois, menus et boutons peuvent être sensiblement modifiés selon l'application en cours, tableur ou **Viewer**).

Retenez aussi que vous pouvez tester et retester tous les calculs que vous voulez, jusqu'à trouver le bon car, à chaque fois,

vous pouvez effacer le dernier calcul réalisé et recommencer autant de fois que vous le voulez. Donc, un autoapprentissage par essai-erreur est tout à fait possible. À ce propos, un truc, le très pratique bouton **[Rappeler toutes les boîtes de dialogue récemment utilisées]** (quatrième bouton en partant de la gauche dans le tableur et sixième bouton dans le **Viewer**), dans la barre d'outil, qui vous permet d'accéder, en deux clics, à l'un des 12 derniers calculs réalisés. À utiliser sans retenue ! (ce que nous ne manquerons pas de faire désormais).

Nous venons de voir comment produire des statistiques descriptives concernant une variable quantitative continue (comme l'âge) par rapport à une ou plusieurs variables qualitatives (comme le groupe d'appartenance et/ou le sexe). Maintenant, en ce qui concerne les autres variables qualitatives qui décrivent votre population (niveau d'études, catégorie socioprofessionnelle, etc.), une autre méthode doit être utilisée pour pouvoir les présenter, il s'agit des « **Tableaux croisés** ».

Procédure SPSS : variable « Niveau d'études »

Cliquez au menu : **Analyse/Statistiques descriptives/Tableaux croisés**, vous obtenez la fenêtre « **Tableaux croisés** ». La façon de procéder est toujours la même (quelques clics de souris) afin de transférer :

a. La variable à décrire (« **Niveau d'études** ») dans la cellule « **Ligne(s) :** ».
b. La variable de regroupement (« **Sexe** ») dans la cellule « **Colonne(s) :** ».
c. La variable de sous-regroupement (« **Groupe** ») dans la cellule « **Couche 1 de 1** » (vous pouvez mettre plusieurs couches, mais la lecture du tableau risque de ne plus être très claire…).

Ce qui donne ce qui suit (Fig. 55) :

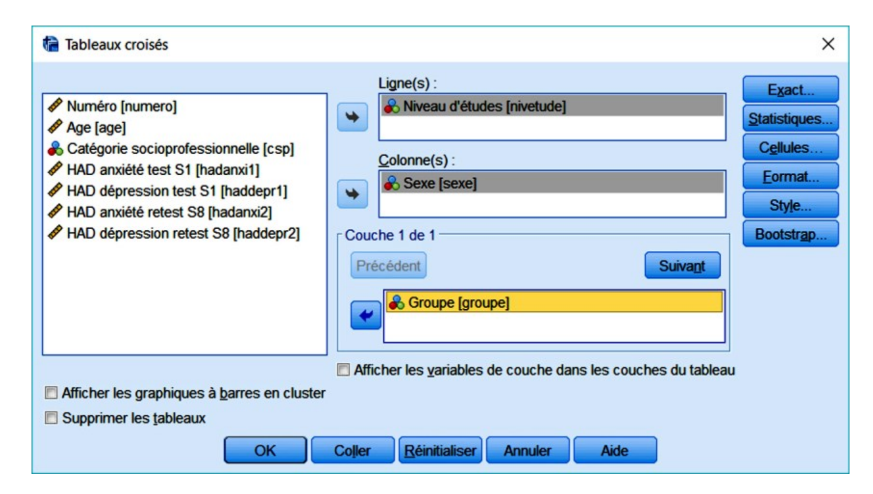


Figure 55 : Paramétrage des tableaux croisés.

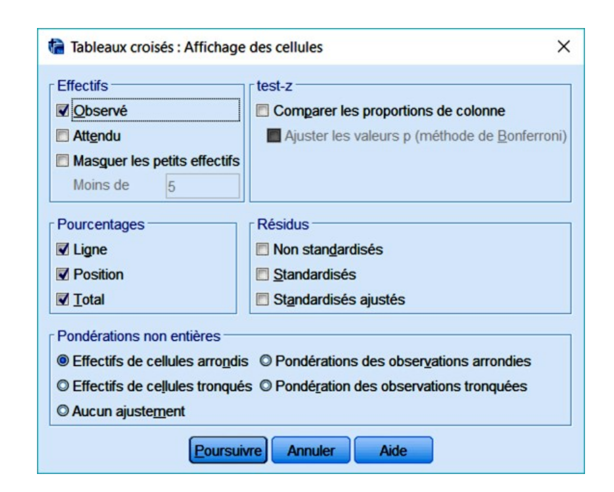


Figure 56 : Paramétrage d'affichage des cellules (pourcentages).

Ensuite, au niveau du bouton **[Statistiques...]**, pour le moment vous ne demandez rien (inutile de le cliquer). Par contre, cliquez sur le bouton **[Cellules...]** pour sélectionner l'affichage des pourcentages par **Ligne**, **Position** et **Total**. Cliquez sur **[Poursuivre]** (Fig. 56).

Tableau croisé Niveau d'études * Sexe * Groupe

Groupe				Sexe		
				Homme	Femme	Total
Expérimental	Niveau d'études	Primaire	Effectif	1	0	1
			% dans Niveau d'études	100,0%	0,0%	100,0%
			% dans Sexe	8,3%	0,0%	5,0%
			% du total	5,0%	0,0%	5,0%
		Collège	Effectif	1	4	5
			% dans Niveau d'études	20,0%	80,0%	100,0%
			% dans Sexe	8,3%	50,0%	25,0%
			% du total	5,0%	20,0%	25,0%
		Lycée	Effectif	10	3	13
			% dans Niveau d'études	76,9%	23,1%	100,0%
			% dans Sexe	83,3%	37,5%	65,0%
			% du total	50,0%	15,0%	65,0%
		Université	Effectif	0	1	1
			% dans Niveau d'études	0,0%	100,0%	100,0%
			% dans Sexe	0,0%	12,5%	5,0%
			% du total	0,0%	5,0%	5,0%
	Total		Effectif	12	8	20
			% dans Niveau d'études	60,0%	40,0%	100,0%
			% dans Sexe	100,0%	100,0%	100,0%
			% du total	60,0%	40,0%	100,0%

Tableau 2 : Résultats du tableau croisé Groupe (Expérimental)/Sexe/Niveau d'études.

Suite du tableau…

Contrôle	Niveau d'études	Primaire	Effectif	1	3	4
			% dans Niveau d'études	25,0%	75,0%	100,0%
			% dans Sexe	12,5%	25,0%	20,0%
			% du total	5,0%	15,0%	20,0%
		Collège	Effectif	3	7	10
			% dans Niveau d'études	30,0%	70,0%	100,0%
			% dans Sexe	37,5%	58,3%	50,0%
			% du total	15,0%	35,0%	50,0%
		Lycée	Effectif	3	1	4
			% dans Niveau d'études	75,0%	25,0%	100,0%
			% dans Sexe	37,5%	8,3%	20,0%
			% du total	15,0%	5,0%	20,0%
		Université	Effectif	1	1	2
			% dans Niveau d'études	50,0%	50,0%	100,0%
			% dans Sexe	12,5%	8,3%	10,0%
			% du total	5,0%	5,0%	10,0%
	Total		Effectif	8	12	20
			% dans Niveau d'études	40,0%	60,0%	100,0%
			% dans Sexe	100,0%	100,0%	100,0%
			% du total	40,0%	60,0%	100,0%

Tableau 3 : Résultats du tableau croisé Groupe (Contrôle)/Sexe/Niveau d'études.

Il ne vous reste plus qu'à cliquer sur **[OK]** pour obtenir ceci (Tab. 2-3, vous pouvez changer le titre par : « **Tableau des caractéristiques Groupe/Sexe/Niveau d'études** »).

Procédure SPSS : variable « Catégorie socioprofessionnelle »

Pour notre dernière variable « **Catégorie socioprofessionnelle** », la procédure est identique. Cliquez maintenant sur le bouton **[Rappeler toutes les boîtes de dialogue…]** et choisissez l'option « **Tableaux croisés** ». Dans la fenêtre, placez « **Catégorie**

socioprofessionnelle » au lieu de « **Niveau d'études** » dans la cellule « **Ligne(s) :** ».

Pour changer, nous n'allons considérer que la variable de regroupement « **Groupe** », donc vous retirez (remettre dans la liste à gauche) de « **Colonne(s) :** » la variable « **Sexe** », de « **Couche 1 de 1** » la variable « **Groupe** ». Vous mettez ensuite dans « **Colonne(s) :** » cette variable « **Groupe** ».

Bouton **[Cellules...]**, vous décochez les pourcentages (pourquoi pas ?!). Bouton **[Poursuivre]**. Bouton **[OK]**, ce qui donne (Tab. 4) :

Tableau croisé Catégorie socioprofessionnelle * Groupe

Effectif

		Groupe		
		Expérimental	Contrôle	Total
Catégorie socioprofessionnelle	Cadre	0	1	1
	Employé	7	9	16
	Ouvrier	9	6	15
	Chômeur	4	4	8
Total		20	20	40

Tableau 4 : Résultats du tableau croisé Catégorie socioprofessionnelle/Groupe

Transferts vers d'autres applications

Les statistiques descriptives de notre étude fictive sont terminées. Vous avez pu constater à la fois la simplicité de SPSS, sa souplesse et la grande liberté de choix, pour faire exactement ce que vous voulez faire. Tout ceci s'applique à tous les calculs que nous aborderons plus loin.

Concernant ces premiers résultats, si vous vous demandez « et après ? », n'oubliez pas que nous sommes dans le cadre d'une étude fictive. Je suppose qu'à partir de vos propres données, vous

saurez apprécier des résultats comme, par exemples, les proportions d'hommes et de femmes selon vos groupes ou encore, selon les niveaux d'études et le groupe. Nous verrons plus loin, dans le chapitre de « **3.6. L'équivalence des groupes** », si ces données ont une signification à prendre en compte ou pas. Ou si, finalement, les groupes sont à peu près équivalents.

Se pose à présent la question de la présentation finale de vos statistiques, c'est-à-dire dans le mémoire ou dans l'article. Ici aussi, le **Viewer** de SPSS vous autorise plusieurs solutions pratiques et intéressantes. Tout d'abord, la première étape, commune à toutes ces solutions, est l'opération classique en informatique qui consiste à sélectionner l'objet que l'on veut copier et à le mettre en mémoire tampon. Par exemple, vous voulez transférer dans un fichier **.doc** (ou **.docx**) de Word (votre mémoire) le tableau intitulé « **Tableau des caractéristiques Groupe/Sexe/Age** », après avoir ouvert l'application Word, vous faites les opérations suivantes :

a. Paramétrage des méthodes de « copier » : Menu du **Viewer Edition/Copier spécial**, cochez les cases (si ce n'est déjà fait) : **Texte normal**, **Rich Text**, **Images**, **Feuille de calcul Excel** et, tout en bas, **Enregistrer par défaut**. Cliquez sur **[OK]**. Cela vous permettra d'avoir plusieurs possibilités de collage dans votre fichier Word.
b. Toujours dans le **Viewer**, ouvert avec « **Statistiques-descriptives.spv** », vous sélectionnez (cliquez sur) le tableau en question (« **Tableau des caractéristiques Groupe/Sexe/Age** »).
c. Ensuite, soit vous allez au menu **Edition/Copier**, soit vous faites un clic droit sur le tableau sélectionné et vous choisissez l'option **« Copier »** du menu contextuel, soit vous appuyez en même temps sur les touches **Ctrl+c**, dans les deux cas vous placez le tableau sélectionné en mémoire tampon.

Vous allez dans votre fichier Word et là vous avez deux

possibilités :

1. Faire **Ctrl+v** (ou bien clic droit et cliquer sur le premier bouton de collage **[Conserver la mise en forme source]**) pour obtenir un tableau Word qui reproduit fidèlement la mise en forme du **Viewer** de SPSS (méthode recommandée).
2. Faire ruban **Accueil**, groupe **Presse-papier**, premier bouton à gauche **[Coller]**, cliquez sur la petite flèche en bas et dans le menu contextuel qui s'ouvre, sélectionnez l'option « **Image (PNG)** » ou « **Image (JPEG)** » pour obtenir un tableau sous la forme d'une image. Son apparence est fidèle au **Viewer**, mais sa résolution laisse fortement à désirer (image de faible qualité, que ce soit en PNG ou en JPEG). Si les contraintes de votre manuscrit vous imposent des fichiers images, plutôt que des tableaux texte, passez plutôt par une copie d'écran du tableau dans Word (possibilité 1) et un logiciel de traitement d'images.[24]

Contrôle de la distribution des fréquences

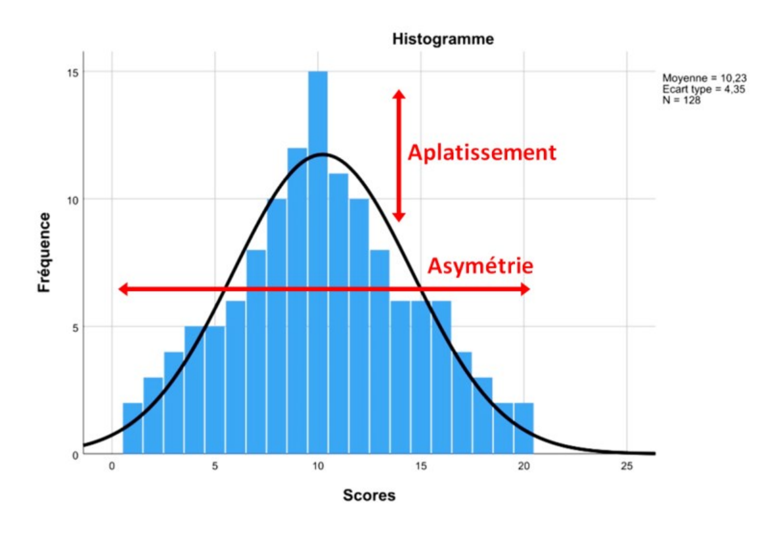


Figure 57 : Exemple de distribution idéalement normale.

[24] Nous pouvons vous recommander Gimp (*https://bit.ly/1SXiBqz*).

Nous verrons qu'un test (l'ANOVA) nécessite une certaine normalité des variables testées. Qu'est-ce que cela signifie ? Une variable est dite normale lorsque la distribution de ses fréquences suit une courbe… normale. Cela veut dire que si l'on considère la fréquence (le nombre de fois) qu'apparaissent chacune des valeurs à cette variable, on doit obtenir une tendance centrale, où se regroupent la plupart des individus de l'échantillon et peu d'individus aux extrémités (valeurs les plus basses et les plus élevées). Aussi, avant de se lancer dans une ANOVA, il vous faut vérifier la normalité de la distribution. Un exemple (presque parfait) ci-dessus (Fig. 57).

Ouvrir dans SPSS le fichier de données : Test-retest-2.sav

Tout d'abord, vous allez ouvrir une nouvelle base de données (créer un nouveau fichier **.sav** de SPSS), le fichier « **Test-retest-2.sav** », de façon à avoir isolé dans des variables différentes les données test et retest pour chacun des groupes. Vous allez à nouveau extraire vos données du fichier Excel « **Data.xlsx** ». Dans SPSS, menu **Fichier/Importer des données/Excel**.

Dans la fenêtre « **Lire le fichier Excel** », vous sélectionnez cette fois la « **Feuille de calcul** » : « **Test-restest 2** ».

En parallèle, à partir de la fenêtre du Viewer, vous allez ouvrir (si ce n'est déjà fait), le fichier de résultats : « **Statistiques-descriptives.spv** », de façon à ce que les résultats concernant les contrôles de distribution s'y affichent à la suite de vos premières statistiques descriptives.

Première opération : libeller les variables. Dans l'onglet **Vue des variables**, vous ajoutez les libellés suivants :

- numero1 : **Numéro Expérimental**
- gredxpe : **Groupe Expérimental**
- hadanxi1 : **HAD Anxiété test Expérimental**

- haddepr1 : **HAD Dépression test Expérimental**
- hadanxi2 : **HAD Anxiété retest Expérimental**
- haddepr2 : **HAD Dépression retest Expérimental**
- numero2 : **Numéro Contrôle**
- grcontr : **Groupe Contrôle**
- hadanxi3 : **HAD Anxiété test Contrôle**
- haddepr3 : **HAD Dépression test Contrôle**
- hadanxi4 : **HAD Anxiété retest Contrôle**
- haddepr4 : **HAD Dépression retest Contrôle**

Concernant la colonne « **Mesure** », n'écoutait pas SPSS et paramétrez toutes les variables des tests en « **Echelle** » et les « variables » de groupes en « **Nominales** » (peu importe, de toute façon elles ne seront pas utilisées).

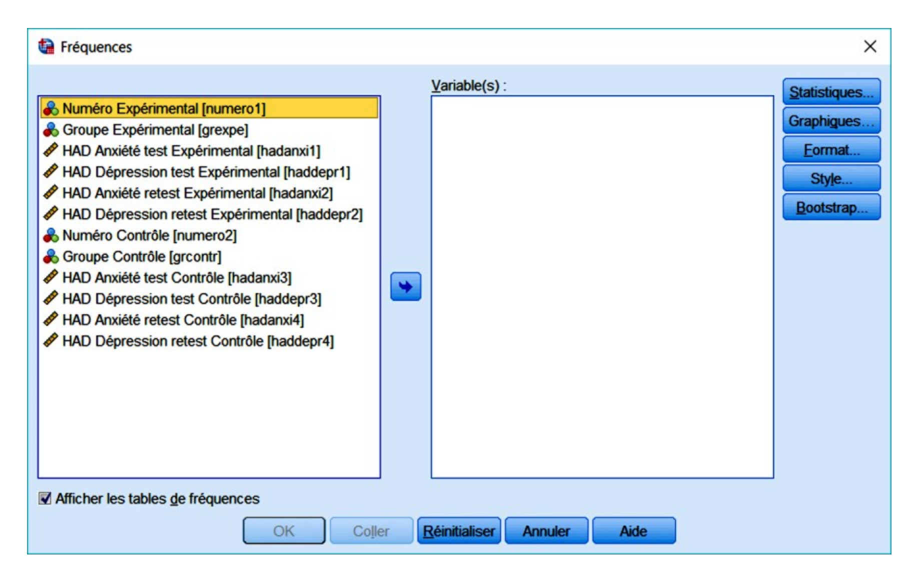


Figure 58 : Fenêtre d'analyse des fréquences.

À présent, pour procéder au contrôle de la distribution des fréquences, vous allez au menu **Analyse/Statistiques**

descriptives/Fréquences, ce qui ouvre la fenêtre ci-dessus (Fig. 58, n'oubliez pas que vous pouvez la redimensionner pour mieux voir les étiquettes des variables).

Glissez toutes les VD (les résultats des tests que vos sujets ont passés) dans la cellule « **Variable(s) :** » (un truc : vous sélectionnez la variable « **HAD Anxiété Test Expérimental** », vous tenez pressée la touche **Maj** du clavier et vous cliquez sur « **HAD Dépression Retest Expérimental** », ce qui sélectionne les 4 variables du groupe expérimental. Il ne vous reste plus qu'à cliquer sur la flèche bleue pour les transférer à droite. Après calcul, vous procéderez pareillement pour les 4 variables du groupe de contrôle. Cela donne (Fig. 59) :

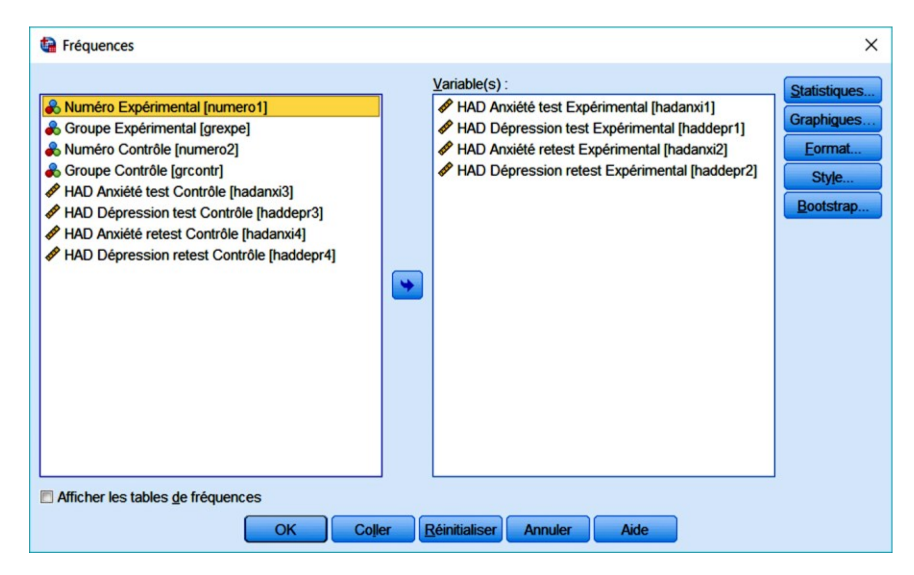


Figure 59 : Fenêtre d'analyse des fréquences paramétrée groupe expérimental.

Bouton **[Statistiques...]**, ne cochez que « **Asymétrie** » et « **Kurtosis** » (aplatissement), bouton **[Poursuivre]**. Pensez à décocher, si non nécessaire, « **Afficher les tables de fréquences** » (à moins que vous ne vouliez vraiment y jeter un coup d'œil). Puis, bouton **[Graphiques...]**, vous sélectionnez dans la petite fenêtre l'option « **Histogrammes :** » et vous cochez « **Avec la**

courbe gaussienne sur l'histogramme » (pour avoir l'histogramme avec le tracé de la courbe normale). Bouton **[Poursuivre]**. C'est prêt, bouton **[OK]**. En plus des histogrammes, le **Viewer** vous affiche les tableaux suivants (Tab. 5-6) :

Statistiques

		HAD Anxiété test Expérimental	HAD Dépression test Expérimental	HAD Anxiété retest Expérimental	HAD Dépression retest Expérimental
N	Valide	20	20	20	20
	Manquant	0	0	0	0
Asymétrie		,482	,025	,511	,836
Erreur standard d'asymétrie		,512	,512	,512	,512
Kurtosis		-,584	-,554	-1,228	,229
Erreur standard de Kurtosis		,992	,992	,992	,992

Tableau 5 : Résultat du test de normalité groupe Expérimental.

Statistiques

		HAD Anxiété test Contrôle	HAD Dépression test Contrôle	HAD Anxiété retest Contrôle	HAD Dépression retest Contrôle
N	Valide	20	20	20	20
	Manquant	0	0	0	0
Asymétrie		-,311	-,060	-1,222	-1,138
Erreur standard d'asymétrie		,512	,512	,512	,512
Kurtosis		-1,077	-,278	1,097	1,127
Erreur standard de Kurtosis		,992	,992	,992	,992

Tableau 6 : Résultat du test de normalité groupe Contrôle.

Certes, on peut voir parfois quelques tendances en regardant un histogramme et une courbe, mais pour être précis, il vaut mieux se référer à deux valeurs, qui caractérisent une distribution : l'asymétrie (*Skewness*, en anglais), c'est-à-dire le fait que la tendance centrale de la courbe (le point le plus haut, qui

correspond à la plus grande fréquence des valeurs) est au centre ou décalée vers la gauche ou la droite. L'aplatissement (*Kurtosis*), c'est-à-dire le fait que la courbe est plus ou moins haute ou basse, pointue ou écrasée. Ces deux caractéristiques ont pour valeur dans le cadre d'une courbe dite « normale » :

Asymétrie (*Skewness*) = 0,000

Aplatissement (*Kurtosis*) = 0,263

Ce qui correspond à une jolie courbe, bien centrée et ni trop haute, ni trop basse. Si l'on reprend le tableau ci-dessus, les caractéristiques de fréquences de nos variables peuvent se résumer ainsi (Tab. 7) :

Groupe	Variable	Symétrie	Aplatissement	Commentaires
Expérimental	Anxiété Test	0.482	-0.584	*Non normale*
	Dépression Test	0.511	-1.228	*Non normale*
	Anxiété Retest	0.025	-0.554	*Seulement symétrie*
	Dépression Retest	0.836	0.229	*Seulement aplatissement*
Contrôle	Anxiété Test	-0.311	-1.077	*Non normale*
	Dépression Test	-1.222	1.097	*Non normale*
	Anxiété Retest	-0.060	-0.278	*Non normale*
	Dépression Retest	-1.138	1.127	*Non normale*

Tableau 7 : Synthèse des résultats aux tests de normalité.

Comme nous sommes dans un contexte de données fictives, il n'est pas utile de s'étendre en commentaires sur l'anormalité de ces distributions.

Concrètement, au niveau de votre mémoire ou article, il est inutile de présenter les histogrammes. Nous ne sommes qu'au stade des variables descriptives. Aussi, les tableaux de données sont bien suffisants. Vous pouvez présenter le tableau de synthèse (Fig. 66) ou bien les tableaux du **Viewer** (recommandé), le lecteur se débrouillera bien.

Après avoir présenté les variables descriptives, c'est-à-dire après avoir décrit en chiffres votre population et l'état de vos variables, vous devez vérifier que vos groupes sont équivalents (ou du moins montrer dans quelle mesure ils ne divergent pas trop), sauf bien entendu, en ce qui concerne la variable indépendante (VI) que vous manipulez (où là, la divergence est voulue et nécessaire).

3.6. L'équivalence des groupes

L'équivalence des groupes est établie en réalisant des tests statistiques entre la variable « **Groupe** » et les autres variables descriptives. Ces tests sont de deux types, selon la nature qualitative ou quantitative de la variable descriptive testée.

1. Tester une variable qualitative

La variable « **Groupe** » étant elle-même une variable qualitative, le test entre ces deux types de variables est un « **Khi-deux** ».[25] Nous vous rappelons ici que la philosophie de ce document est de faire simple et « **classique** », aussi, même si d'autres tests existent, nous nous garderons bien de vous encombrer l'esprit en vous en parlant.

---

[25] $\chi 2$ – la dénomination normale devrait être « Khi-puissance deux » ou « Khi-carré ». Mais pour simplifier tout le monde dit « Khi-deux ».

Procédure SPSS : test de Khi-deux Groupe-Sexe-Niveau d'études-Catégorie socioprofessionnelle

Ouvrir dans SPSS le fichier de données : Principale.sav

Au départ, la procédure est identique à celle de la présentation des variables descriptives qualitatives, il s'agit d'un **Tableau croisé**. Cliquez au menu : **Analyse/Statistiques descriptives/Tableaux croisés** ou, plus rapidement, sélectionnez « **Tableaux croisés** » après avoir cliqué sur le bouton de rappel des boîtes de dialogue.

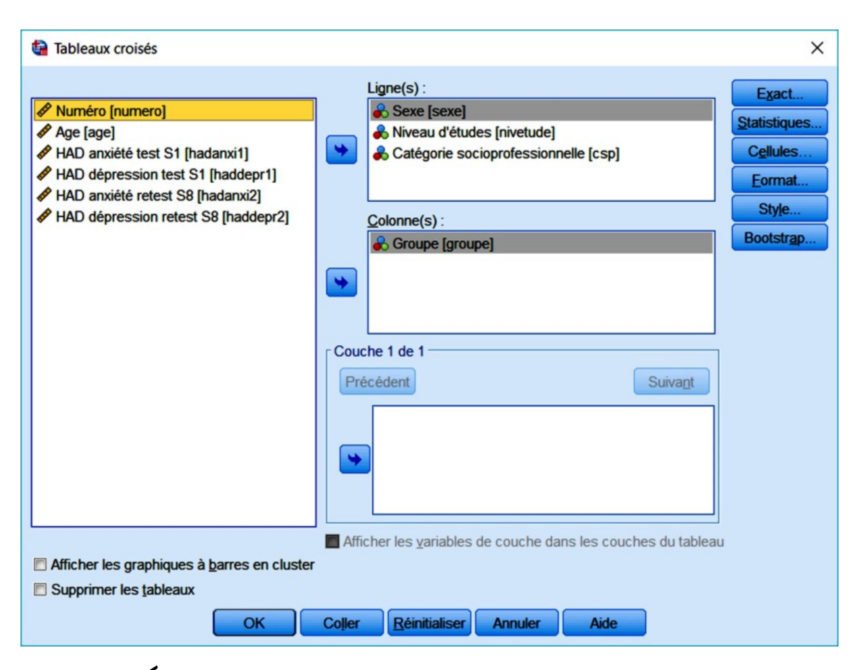


Figure 60 : Équivalence des groupes sur les variables qualitatives.

Glissez la variable « **Groupe** » dans la cellule « **Colonne(s) :** ». Cette fois, nous allons voir comment SPSS peut être très rapide, une fois que l'on a bien compris comment ça marche. Au lieu de traiter les variables à tester une par une, il vous est tout à fait possible de faire les choses d'un seul coup. Aussi, vous allez glisser dans la cellule « **Ligne(s) :** » les trois variables qualitatives à tester : « **Sexe** », « **Niveau d'études** » et « **Catégorie**

socioprofessionnelle » (pour sélectionner plusieurs variables qui ne se suivent pas dans la liste, maintenez enfoncée la touche **<Ctrl>** du clavier et sélectionnez à la souris chaque variable), ce qui donne (Fig. 60).

Ensuite, cliquez sur le bouton **[Statistiques...]**, dans la petite fenêtre cochez l'option « **Khi-carré** », bouton **[Poursuivre]**. Éventuellement, faites un petit tour par le bouton **[Cellules...]**, pour vérifier que les pourcentages sont bien décochés, car ils ne feraient qu'encombrer l'affichage. Il ne reste plus qu'à cliquer **[OK]**. Vous obtenez un rappel descriptif des variables (les effectifs), utile pour pouvoir réfléchir, et les Khi-deux, pour le sexe (Tab. 8).

Sexe * Groupe

Tests du khi-carré

	Valeur	ddl	Signification asymptotique (bilatérale)	Sig. exacte (bilatérale)	Sig. exacte (unilatérale)
khi-carré de Pearson	1,600[a]	1	,206		
Correction pour continuité[b]	,900	1	,343		
Rapport de vraisemblance	1,611	1	,204		
Test exact de Fisher				,343	,172
Association linéaire par linéaire	1,560	1	,212		
N d'observations valides	40				

a. 0 cellules (,0%) ont un effectif théorique inférieur à 5. L'effectif théorique minimum est de 10,00.

b. Calculée uniquement pour une table 2x2

Tableau 8 : Résultat de l'équivalence des groupes sur le sexe.

Pour le niveau d'études (Tab. 9) :

Niveau d'études * Groupe

Tests du khi-carré

	Valeur	ddl	Signification asymptotique (bilatérale)
khi-carré de Pearson	8,565[a]	3	,036
Rapport de vraisemblance	8,983	3	,030
Association linéaire par linéaire	3,764	1	,052
N d'observations valides	40		

a. 4 cellules (50,0%) ont un effectif théorique inférieur à 5. L'effectif théorique minimum est de 1,50.

Tableau 9 : Résultat de l'équivalence des groupes sur le niveau d'études.

Pour la catégorie socioprofessionnelle (Tab. 10) :

Catégorie socioprofessionnelle * Groupe

Tests du khi-carré

	Valeur	ddl	Signification asymptotique (bilatérale)
khi-carré de Pearson	1,850[a]	3	,604
Rapport de vraisemblance	2,241	3	,524
Association linéaire par linéaire	,075	1	,784
N d'observations valides	40		

a. 4 cellules (50,0%) ont un effectif théorique inférieur à 5. L'effectif théorique minimum est de ,50.

Tableau 10 : Résultat de l'équivalence des groupes sur la catégorie socioprofessionnelle.

À part le « **Khi-carré de Pearson** », vous notez qu'il y a d'autres informations et quelques petites différences. Cela est dû aux caractéristiques propres à chaque variable, mais dans le cadre de ce manuel, nous ne retiendrons que le « **Khi-carré de Pearson** », ainsi que sa « **Signification asymptotique** » (donc la première ligne dans chaque tableau). Si l'on fait une synthèse, le contrôle

de l'équivalence des groupes « **Expérimental** » et « **Contrôle** », donne le tableau suivant (Tab. 11) :

Variables	**χ^2 et Signification**
Sexe	χ^2 = 1,600 ; *P* = 0,206
Niveau d'études	χ^2 = 8,565 ; *P* = 0,036
Catégorie socioprofessionnelle	χ^2 = 1,850 ; *P* = 0,604

Tableau 11 : Synthèse des Khi-carré pour l'équivalence des groupes.

Interprétation du Khi-deux (χ2)

Cette interprétation est très simple. Plus le **χ2** est élevé et plus la différence entre les groupes est significative. Mais comme cette valeur du **χ2** a un côté un peu « relatif », il est bon de considérer en même temps (et nous avons envie de dire surtout) la « **Signification asymptotique (bilatérale)** », c'est-à-dire le coefficient de signification du **χ2**, que l'on écrit aussi ***P***. Pour ***P***, c'est l'inverse, plus il est petit et plus la différence entre groupes est significative pour la variable considérée. Retenez qu'habituellement on utilise un seuil de **0,05** pour ***P***. Cela signifie que l'on considère une différence significative pour ***P* <= 0,05**, la différence n'étant pas significative pour ***P* > 0,05**. Forts de ces précisions, examinons nos **χ2**.

Nous voyons qu'il n'y a pas de différence significative entre les groupes pour le sexe (**χ2 = 1,600 ; *P* = 0,206**) et pour la catégorie socioprofessionnelle (**χ2 = 1,850 ; *P* = 0,604**), dans les deux cas ***P*** est supérieur à ***0,05***. L'examen des effectifs montre qu'effectivement, les effectifs sont à peu près semblables entre les deux groupes.

Par contre, il y a une différence significative entre les groupes en ce qui concerne le niveau d'études (**χ2 = 8,565 ; *P* = 0,036**), ***P*** est ici bien inférieur à **0,05**. Lorsqu'on souhaite une équivalence,

constater une différence est gênant, car cela signifie que la différence attendue pour nos VD au retest (ici anxiété et dépression) pourrait être due à ce fameux « **niveau d'études** » et non pas tant au suivi psychologique (en tout cas il serait difficile de distinguer les deux influences). Cependant, souvent un examen des chiffres permet de relativiser cet effet.

On voit bien au niveau des effectifs (voir le tableau descriptif dans le fichier des résultats, dans le **Viewer**) que la différence est surtout due au fait que le groupe « **Expérimental** » a **5** « **Collège** » et **13** « **Lycée** », tandis que le groupe de « **Contrôle** » a les effectifs inverses, **10** « **Collège** » et **4** « **Lycée** ». Si la différence s'était jouée aux deux extrémités des niveaux d'études (par exemple, un nombre considérable de niveau primaire pour le groupe « **Expérimental** » et un nombre tout aussi considérable de niveau universitaire pour le groupe de « **Contrôle** »), nous pourrions nous affoler. Là, entre collège et lycée, on peut considérer cette variable, même si elle différencie les groupes, comme négligeable sur les niveaux d'anxiété et de dépression, liés à la présence ou non de suivi psychologique (on pourrait aussi neutraliser cette différence en remplaçant les catégories « Collège » et « Lycée » par « Enseignement secondaire »).

Nota : la syntaxe entre parenthèses, par exemple, de « **($\chi2$ = 1,600 ; *P* = 0,206)** » est celle que vous devez utiliser dans votre mémoire ou article. Exemple de phrase :

« On ne trouve pas de différence significative entre les groupes pour le sexe **($\chi2$ = 1,600 ; *P* = 0,206)** et pour la catégorie socioprofessionnelle **($\chi2$ = 1,850 ; *P* = 0,604)**. »

2. Tester une variable quantitative

Pour tester une variable quantitative, continue, comme l'âge, face à une variable qualitative, comme le groupe, on emploie une analyse de variance (ou ANOVA, de l'anglais *ANalysis Of VAriance*). Dans le cadre de notre recherche fictive, nous devons

tester l'équivalence des groupes en ce qui concerne l'âge, mais aussi en ce qui concerne l'état psychologique (anxiété et dépression) au test, à S1. Car un effet du suivi psychologique sur anxiété et dépression ne pourrait être affirmé si déjà, au départ, un groupe est plus ou moins anxieux ou dépressif qu'un autre.

Procédure SPSS : Test ANOVA Groupe-Age-HAD anxiété et dépression

Ouvrir dans SPSS le fichier de données : Principale.sav

Au menu allez à **Analyse/Comparer les moyennes/ANOVA à 1 facteur**. Glissez dans la cellule « **Liste Variables dépendantes :** » (on ne peut pas plus explicite) les variables quantitatives qui nous intéressent : « **Age** », « **HAD anxiété test S1** » et « **HAD dépression test S1** ». Puis, glissez la variable « **Groupe** » dans la cellule « **Facteur :** », ce qui donne (Fig. 61) :

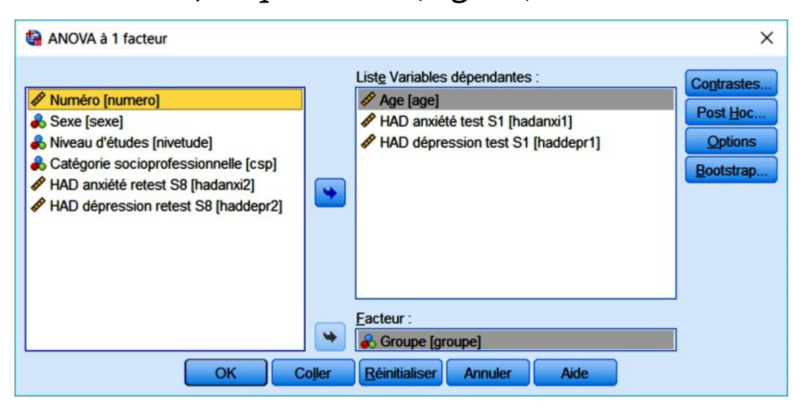


Figure 61 : Test ANOVA sur les variables quantitatives continues selon le groupe.

Bouton **[Options]**, sélectionnez « **Caractéristique** », puis, bouton **[Poursuivre]**. De retour dans la fenêtre de base de l'ANOVA, cliquez sur le bouton [OK] pour lancer le calcul des trois ANOVA en même temps. Le résultat donne (Tab. 12) :

ANOVA

		Somme des carrés	ddl	Carré moyen	F	Sig.
Age	Intergroupes	2,025	1	2,025	,024	,877
	Intragroupes	3166,750	38	83,336		
	Total	3168,775	39			
HAD anxiété test S1	Intergroupes	2,500	1	2,500	,446	,508
	Intragroupes	213,100	38	5,608		
	Total	215,600	39			
HAD dépression test S1	Intergroupes	3,025	1	3,025	,376	,543
	Intragroupes	305,350	38	8,036		
	Total	308,375	39			

Tableau 12 : Résultat du test ANOVA sur les variables quantitatives continues selon le groupe.

Interprétation de l'ANOVA

L'ANOVA s'interprète à partir des valeurs de « **F** » et de « **Sig.** » (Signification) dans le tableau de résultats. Tout comme le **χ2**, plus **F** est grand et plus la différence entre les groupes est significative. En parallèle, plus la signification (***P***) est petite et plus la différence entre groupes est significative. Cela signifie, comme au **χ2**, que l'on considère une différence significative pour ***P*** **<= 0,05**, la différence n'étant pas significative pour ***P*** **> 0,05**. Les tests ANOVA donnent (Tab. 13) :

Variables	**ANOVA : F et Signification**
Age	F(1,38) = 0,024 ; *P* = 0,877
HAD anxiété test S1	F(1,38) = 0,446 ; *P* = 0,508
HAD dépression test S1	F(1,38) = 0,376 ; *P* = 0,543

Tableau 13 : Synthèse du test ANOVA sur les variables quantitatives continues selon le groupe.

Nous constatons qu'il n'y a pas de différence significative entre les groupes (ce qui est confirmé en regardant les moyennes au tableau des statistiques descriptives), en ce qui concerne l'« **Age** » (**F(1,38) = 0,024 ; *P* = 0,877**), la « **HAD anxiété test S1** » (**F(1,38) = 0,446 ; *P* = 0,508**) et la « **HAD dépression test S1** » (**F(1,38) = 0,376 ; *P* = 0,543**), tous les **F** sont faibles, inférieurs à zéro et les ***P*** bien supérieurs à **0,05**. Les groupes sont donc bien équivalents pour ces variables à S1.

Notez que la formule syntaxique, par exemple, de (**F(1,38) = 0,024 ; *P* = 0,877**) est celle que vous devez utiliser pour parler des tests ANOVA dans votre mémoire ou article. Le « **(1,38)** » entre parenthèses n'est pas un nombre réel, mais l'indication des degrés de liberté (ddl) de l'ANOVA : **1** entre les groupes (Inter-groupes) et **38** dans les groupes (Intra-groupes).

3.7. Effets des variables descriptives

Maintenant que nous savons que les groupes sont à peu près équivalents, avant de voir l'effet groupe au retest à S8 (les staistiques inférentielles) nous devons nous assurer que les autres variables descriptives n'ont pas déjà un effet sur les VD au test à S1. En ce qui concerne les variables qualitatives, on considère que chaque variable produit des sous-groupes (autant qu'il y a de niveaux à la variable considérée, c'est-à-dire 2 pour le sexe, 4 pour le niveau d'études et la CSP). Il suffit alors de comparer ces sous-groupes avec une ANOVA.

La procédure SPSS reprend celle précédemment décrite pour l'équivalence des groupes. Dans la fenêtre « **ANOVA à 1 facteur** » on place dans la cellule « **Liste Variables dépendantes :** » les deux VD à S1 « HAD anxiété test S1 » et « HAD dépression test S1 ». Et dans la cellule « **Facteur :** », on placera tour à tour les variables « Sexe », « Niveau d'études » et « Catégorie socioprofessionnelle ». Bouton **[Options]**, pensez à cocher « **Caractéristique** ». Les résultats sont présentés ci-après (Tab. 14-15-16).

Pour le sexe :

ANOVA

		Somme des carrés	ddl	Carré moyen	F	Sig.
HAD anxiété test S1	Intergroupes	,100	1	,100	,018	,895
	Intragroupes	215,500	38	5,671		
	Total	215,600	39			
HAD dépression test S1	Intergroupes	1,225	1	1,225	,152	,699
	Intragroupes	307,150	38	8,083		
	Total	308,375	39			

Tableau 14 : Résultats pour les effets de la variable descriptive Sexe.

Pour le niveau d'études :

ANOVA

		Somme des carrés	ddl	Carré moyen	F	Sig.
HAD anxiété test S1	Intergroupes	8,729	3	2,910	,506	,680
	Intragroupes	206,871	36	5,746		
	Total	215,600	39			
HAD dépression test S1	Intergroupes	2,046	3	,682	,080	,970
	Intragroupes	306,329	36	8,509		
	Total	308,375	39			

Tableau 15 : Résultats pour les effets de la variable descriptive Niveau d'études.

Pour la catégorie socioprofessionnelle :

ANOVA

		Somme des carrés	ddl	Carré moyen	F	Sig.
HAD anxiété test S1	Intergroupes	9,929	3	3,310	,579	,632
	Intragroupes	205,671	36	5,713		
	Total	215,600	39			
HAD dépression test S1	Intergroupes	15,837	3	5,279	,650	,588
	Intragroupes	292,538	36	8,126		
	Total	308,375	39			

Tableau 16 : Résultats pour les effets de la variable descriptive Catégorie socioprofessionnelle.

Finalement, on constate qu'il n'y a aucune différence significative en ce qui concerne anxiété et dépression à la HAD, que la population soit regroupée selon le « **Sexe** » (anxiété : **(F(1,38) = 0,018 ; *P* = 0,895)** – dépression : **(F(1,38) = 0,152 ; *P* = 0,699))**, selon le « **Niveau d'études** » (anxiété : **(F(1,38) = 0,506 ; *P* = 0,680)** – dépression : **(F(1,38) = 0,080 ; *P* = 0,970))** ou selon la « **Catégorie socioprofessionnelle** » (anxiété : **(F(1,38) = 0,579 ; *P* = 0,632)** – dépression : **(F(1,38) = 0,650 ; *P* = 0,588))**.

Ainsi, « Sexe », « Niveau d'études » et « Catégorie socioprofessionnelle » n'ont pas d'influence sur les VD (anxiété et dépression) au test à S1.

Corrélations avec l'âge

L'âge est une variable quantitative continue, les scores d'anxiété et de dépression sont aussi des variables quantitatives et continues. Dans ce cas, le test de comparaison n'est pas une ANOVA, mais un coefficient de corrélation de Pearson (vous savez la fameuse corrélation entre l'âge et le poids…). La procédure SPSS est la suivante. Rendez-vous au menu **Analyse/Corrélation/Bivariée** (toujours à partir de « **Principale.sav** »).

Il suffit alors de glisser les variables quantitatives à corréler (« **Age** », « **HAD anxiété test S1** », « **HAD dépression test S1** ») dans la cellule « **Variables :** », ce qui donne (Fig. 62) :

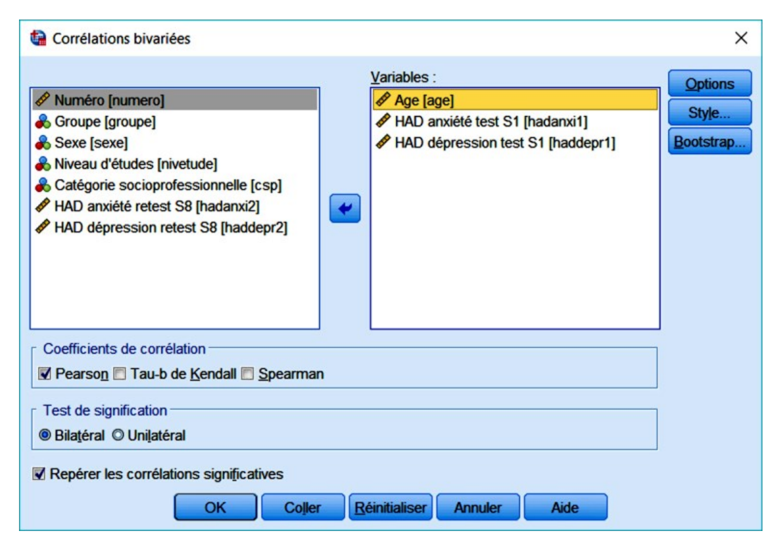


Figure 62 : Corrélations entre l'âge et le test à la HAD.

Corrélations

		Age	HAD anxiété test S1	HAD dépression test S1
Age	Corrélation de Pearson	1	,137	,092
	Sig. (bilatérale)		,399	,571
	N	40	40	40
HAD anxiété test S1	Corrélation de Pearson	,137	1	,448**
	Sig. (bilatérale)	,399		,004
	N	40	40	40
HAD dépression test S1	Corrélation de Pearson	,092	,448**	1
	Sig. (bilatérale)	,571	,004	
	N	40	40	40

**. La corrélation est significative au niveau 0.01 (bilatéral).

Tableau 17 : Résultats des corrélations entre l'âge et le test à la HAD.

Cliquez sur le bouton **[Options]** et sélectionnez « **Moyennes et écarts-types** ». Bouton **[Poursuivre]** et bouton **[OK]**, le résultat est le suivant (Tab. 17).

SPSS affiche dans le **Viewer** une table des corrélations qui présente, au croisement entre chaque ligne et colonne, le coefficient de corrélation de Pearson, ainsi que le niveau de signification, pour les deux variables concernées (attention, tableau à double entrée, les résultats sont affichés deux fois).

Interprétation du coefficient de corrélation

La valeur d'un coefficient de corrélation se situe entre -1 et +1. En théorie, à 0 il n'y a aucune corrélation. Tandis qu'à -1 ou +1 il y a une parfaite corrélation (une corrélation négative signifie simplement que la relation est inversée, par exemple : si vous trouvez une corrélation de +0,55 entre âge et nombre de cheveux blancs, vous trouverez aussi bien une corrélation de -0,55 entre âge et cheveux pas blancs – exemple fictif !).

En sciences humaines, on considère généralement comme significatif un coefficient de corrélation de ± 0,30 et au-dessus, 70 % ou plus de la variance expliquant la relation. Le coefficient de corrélation **r** est généralement suivi de son indice de signification ***P*** qui, comme vous le savez à présent, plus il est petit (**<= 0,05**) et plus il y a corrélation.

Les résultats observés pour nos données fictives sont : entre « **Age** » et « **Anxiété** » il n'y a pas de corrélation **(r = 0,137 ; *P* = 0,399)**, non plus entre « **Age** » et « **Dépression** » **(r = 0,092 ; *P* = 0,571)**. Donc la variable descriptive « **Age** » n'a pas d'influence sur les VD à S1.

Remarquez qu'il y a une significative corrélation entre « **Anxiété** » et « **Dépression** » **(r = 0,448 ; *P* = 0,004)**. Il est normal que ces deux VD soient liées, puisque toutes les deux ne le sont pas avec « **Age** ». Notez que la syntaxe, par exemple **(r = 0,092 ; *P* = 0,571)**, est celle que vous devez employer dans vos écrits. Vous pouvez aussi préciser : **(r = 0,092 ; *P*(bilatéral) = 0,571)**.

Remarque importante avant de poursuivre

Nous venons de voir comment vérifier l'équivalence des groupes et la non-influence des variables descriptives sur les VD. Dans l'idéal, pour que les statistiques inférentielles (celles qui sont destinées à valider ou pas les hypothèses de l'étude) soient valides, il faut que les groupes soient équivalents (le fameux **« toutes choses égales par ailleurs »**), c'est-à-dire qu'il n'y ait pas, au départ, de différences telles entre les groupes que ces différences pourraient expliquer à elles seules les résultats aux VD sans faire appel à l'hypothèse de travail. Il faut aussi que les variables descriptives n'aient aucun effet sur les VD, car sinon, là encore, les résultats se trouveront entachés.

Aussi, que se passe-t-il pour le chercheur lorsqu'il ne peut obtenir ni une « équivalence des groupes » ni une neutralité des variables descriptives ? Eh bien, si vous êtes dans ce cas, inutile de perdre davantage votre temps, votre recherche est compromise et il vous faut recommencer sur une autre population, avec une meilleure sélection et répartition de vos sujets dans les groupes.

En pratique, pour un mémoire universitaire, avec une limite de temps, le mieux est de poursuivre malgré tout votre travail, tout en précisant bien ses faiblesses en matière de statistiques, car ce qu'on vous demande, en tant qu'étudiant, n'est pas tant de valider une hypothèse, que de montrer votre savoir-faire, en tant qu'exercice. Ce qui ne serait pas le cas dans le cadre d'une publication scientifique où là vous devez respecter des principes scientifiques et mettre en œuvre une méthode de recherche rigoureuse.

Nota : En pratique, pour ce qui concerne l'affichage des résultats dans le Viewer de SPSS, vous avez le choix entre afficher ces résultats dans le même fichier **.spv**, ou bien, de créer autant de fichiers **.spv** que vous le désirez. Par exemple, vous pourriez avoir trois fichiers : « **Statistiques descriptives.spv** » ; « **Équivalence des groupes.spv** » ; et « **Effet des variables descriptives.spv** ». En tout cas, pour bien différencier les

résultats, nous vous conseillons, pour ce qui concerne les statistiques inférentielles qui viennent, de placer leurs résultats dans un fichier à part : « **Statistiques inférentielles.spv** ».

3.8. Les statistiques inférentielles

En résumé, après avoir accompli les opérations suivantes :

- décrit la population (statistiques descriptives) ;
- vérifié que les groupes (la variable indépendante et manipulée) sont équivalents ;
- vérifié que les variables descriptives n'ont pas d'effet sur les VD ;

nous pouvons ENFIN (!) faire des statistiques inférentielles, c'est-à-dire vérifier si notre hypothèse, que le suivi psychologique a un effet au bout de 8 semaines de traitement, sur l'anxiété et la dépression (mesurées avec la HAD), est validée ou non.

Comparaison des groupes au retest à S8

Cette comparaison se fait de la même façon que la comparaison des groupes à la HAD au test à S1, qui était destinée à vérifier leur équivalence. Nous avons une variable qualitative « Groupe » et deux variables quantitatives « HAD anxiété » et « HAD dépression », la méthode à employer est donc… une ANOVA.

Ouvrir dans SPSS le fichier de données : « Principale.sav »

La procédure est la même que précédemment : cliquez sur le bouton de rappel des boîtes de dialogue et choisissez l'option « **ANOVA à 1 facteur** ». Dans la cellule « **Liste Variables**

dépendantes : » glissez les deux variables dépendantes à S8 (« **HAD anxiété retest S8** » et « **HAD dépression retest S8** ») et dans la cellule « **Facteur :** » glissez le « **Groupe** ». Bouton **[Options]** cochez « **Caractéristiques** », bouton **[Poursuivre]** (Fig. 63).

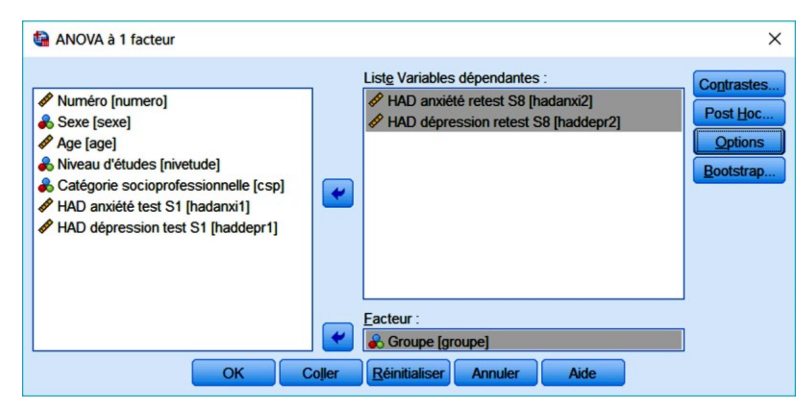


Figure 63 : Paramétrage de l'ANOVA au retest à S8.

Bouton **[OK]**, les résultats s'affichent (Tab. 18) :

ANOVA

		Somme des carrés	ddl	Carré moyen	F	Sig.
HAD anxiété retest S8	Intergroupes	1334,025	1	1334,025	187,509	,000
	Intragroupes	270,350	38	7,114		
	Total	1604,375	39			
HAD dépression retest S8	Intergroupes	950,625	1	950,625	231,934	,000
	Intragroupes	155,750	38	4,099		
	Total	1106,375	39			

Tableau 18 : Résultats de l'ANOVA au retest à S8.

Nous constatons une différence très significative entre le groupe « Expérimental » qui a eu un suivi psychologique durant les 8 semaines et le groupe de « Contrôle » qui n'a pas eu un tel

suivi, en ce qui concerne l'anxiété **(F(1,38) = 187,509 ; *P* = 0,000)** et la dépression **(F(1,38) = 231,934 ; *P* = 0,000)**. Bien entendu, l'hypothèse fictive de cette étude fictive est confirmée, le suivi psychologique a un effet sur l'anxiété et la dépression (dans la réalité, il est rare qu'un effet se produise avec autant d'évidence…).

Comparaison des groupes à eux-mêmes

S'il est possible de comparer les groupes entre eux à S8, on peut affiner le résultat en comparant les groupes à eux-mêmes entre S1 et S8, ceci afin de bien montrer leur évolution. Pour cela, ouvrez dans SPSS le fichier « **Test-retest-1.sav** ».

Rappel : ici nous avons deux variables groupe indépendantes (« **grexpe** » pour le groupe « **Expérimental** » et « **grcontr** » pour le groupe de « **Contrôle** »). Chaque groupe a ses deux colonnes de données test-retest, pour l'anxiété et pour la dépression. Comme nous nous retrouvons dans la situation d'une variable qualitative déterminant des variables quantitatives, la méthode est toujours une ANOVA.

Ouvrir dans SPSS le fichier de données : « Test-retest-1.sav »

Si, pour vous entraîner, vous voulez reconstituer « **Test-restest 1.sav** » à partir de la feuille Excel, pensez à donner un libellé aux variables et à régler leur « **Mesure** » (« **Echelle** » partout, sauf pour les variables groupe qui sont « **Nominales** »).

Pour la comparaison des groupes à eux-mêmes, la procédure est habituelle : cliquez sur le bouton de rappel des boîtes de dialogue et choisissez l'option « **ANOVA à 1 facteur** ». Dans la cellule « **Liste Variables dépendantes :** », glissez les deux variables dépendantes à S8 (« **Anxiété test-retest Expérimental** » et « **Dépression test-retest Expérimental** ») et dans la cellule « **Facteur :** » glissez le groupe « **Expérimental test-retest** ». Bouton **[Options]** cochez « **Caractéristiques** », bouton **[Poursuivre]** (Fig. 64).

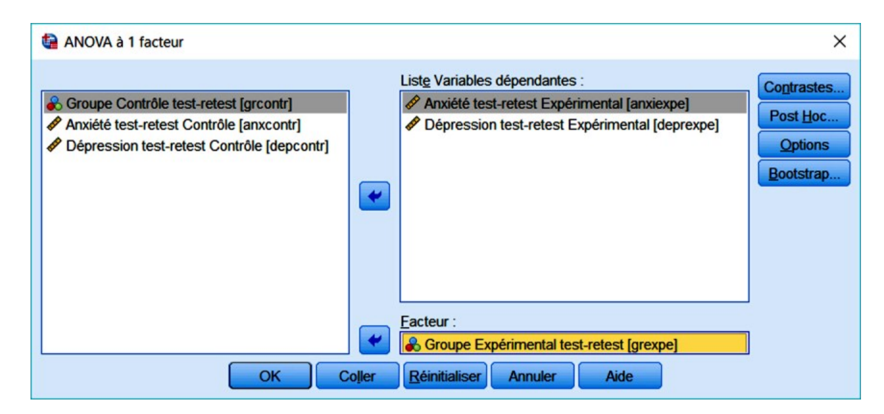


Figure 64 : Paramétrage de l'ANOVA comparant le groupe expérimental à lui-même.

Bouton **[OK]**, les résultats pour le groupe « **Expérimental** » s'affichent (Tab. 19).

ANOVA

		Somme des carrés	ddl	Carré moyen	F	Sig.
Anxiété test-retest Expérimental	Intergroupes	1334,025	1	1334,025	174,593	,000
	Intragroupes	290,350	38	7,641		
	Total	1624,375	39			
Dépression test-retest Expérimental	Intergroupes	837,225	1	837,225	140,431	,000
	Intragroupes	226,550	38	5,962		
	Total	1063,775	39			

Tableau 19 : Résultats de l'ANOVA comparant le groupe expérimental à lui-même.

Même procédure en changeant ce qui doit être changé, pour le groupe Contrôle (Fig. 65 ; Tab. 20).

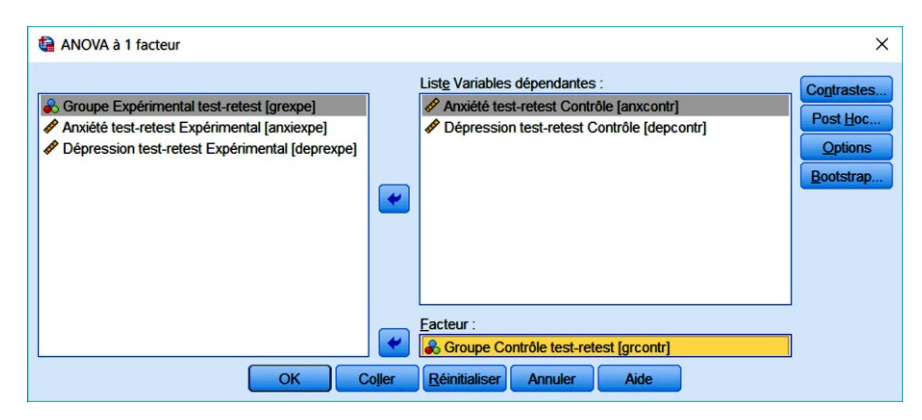


Figure 65 : Paramétrage de l'ANOVA comparant le groupe contrôle à lui-même.

ANOVA

		Somme des carrés	ddl	Carré moyen	F	Sig.
Anxiété test-retest Contrôle	Intergroupes	2,500	1	2,500	,492	,487
	Intragroupes	193,100	38	5,082		
	Total	195,600	39			
Dépression test-retest Contrôle	Intergroupes	13,225	1	13,225	2,143	,151
	Intragroupes	234,550	38	6,172		
	Total	247,775	39			

Tableau 20 : Résultats de l'ANOVA comparant le groupe contrôle à lui-même.

Cette comparaison test-retest montre une différence significative du point de vue de l'anxiété **(F(1,38) = 174,593 ; *P* = 0,000)** et dépression **(F(1,38) = 140,431 ; *P* = 0,000)** pour le groupe « **Expérimental** » qui a eu un suivi psychologique ; alors que le groupe de « **Contrôle** » n'a pas eu d'évolution : anxiété **(F(1,38) = 0,492 ; *P* = 0,487)**, dépression **(F(1,38) = 2,143 ; *P* = 0,151)**.

Autres tests pour petits échantillons

En principe, une ANOVA est plutôt réservée à des échantillons supérieurs à 30 (N > 30). D'ailleurs, dans le cadre de notre étude fictive, nous voyons qu'avec N = 20 pour chaque groupe, cette condition n'est pas remplie. De plus, l'ANOVA demande aussi à ce que les différentes valeurs de la variable se répartissent sur une courbe de distribution normale. Là encore, nous avons vu aux **Statistiques descriptives** que cette condition n'est pas non plus remplie. Lorsqu'il y a petits effectifs et distribution non normale (ce qui est le lot des petits effectifs), deux autres tests sont possibles : le « **test-T de Student** » et le « **test de Wilcoxon** » (qui est un test non-paramétrique, donc réservé pour les petits effectifs).

Tout d'abord, il vous faut rouvrir la base de données, le fichier « **Test-retest-2.sav** », de façon à pouvoir isoler dans des variables différentes les données test et retest pour chacun des groupes.

Ouvrir dans SPSS le fichier de données : « Test-retest-2.sav »

Ici, point de variable de groupe, donc, car nous allons simplement comparer des séries de données deux par deux et, finalement, le choix des groupes se fera au travers du choix des variables. Nous comparerons :

- **Pour le groupe « Expérimental » :** « Anxiété Test Expérimental » à « Anxiété Retest Expérimental » ;
- **Pour le groupe « Expérimental » :** « Dépression Test Expérimental » à « Dépression Retest Expérimental » ;
- **Pour le groupe « Contrôle » :** « Anxiété Test Contrôle » à « Anxiété Retest Contrôle » ;
- **Pour le groupe « Contrôle » :** « Dépression Test Contrôle » à « Dépression Retest Contrôle ».

Procédure pour le test T de Student

Rappel des variables en jeu dans le tableur à partir du fichier de données « **Test-restest 2.sav** » (Fig. 66) :

Test-retest-2.sav [Jeu_de_données1] - IBM SPSS Statistics Editeur de données

Fichier Edition Affichage Données Transformer Analyse Graphiques Utilitaires Extensions Fenêtre Aide

	Nom	Type	Largeur	Décimales	Libellé	Valeurs	Ma
1	numero1	Numérique	2	0	Numéro Expérimental	Aucun	Aucı
2	grexpe	Numérique	1	0	Groupe Expérimental	Aucun	Aucı
3	hadanxi1	Numérique	2	0	HAD Anxiété test Expérimental	Aucun	Aucı
4	haddepr1	Numérique	2	0	HAD Dépression test Expérimental	Aucun	Aucı
5	hadanxi2	Numérique	2	0	HAD Anxiété retest Expérimental	Aucun	Aucı
6	haddepr2	Numérique	2	0	HAD Dépression retest Expérimental	Aucun	Aucı
7	numero2	Numérique	2	0	Numéro Contrôle	Aucun	Aucı
8	grcontr	Numérique	1	0	Groupe Contrôle	Aucun	Aucı
9	hadanxi3	Numérique	2	0	HAD Anxiété test Contrôle	Aucun	Aucı
10	haddepr3	Numérique	2	0	HAD Dépression test Contrôle	Aucun	Aucı
11	hadanxi4	Numérique	2	0	HAD Anxiété retest Contrôle	Aucun	Aucı
12	haddepr4	Numérique	2	0	HAD Dépression retest Contrôle	Aucun	Aucı
13							

Figure 66 : Données pour les tests sur de petits échantillons.

Allez au menu **Analyse/Comparer les moyennes/Test-T pour échantillons appariés** (**remarque :** lorsqu'on fait une comparaison des moyennes, le test-T pour échantillons appariés est aussi utilisé, comme ici, dans le cas d'un même échantillon avec test et retest). La fenêtre obtenue est la suivante (Fig. 67) :

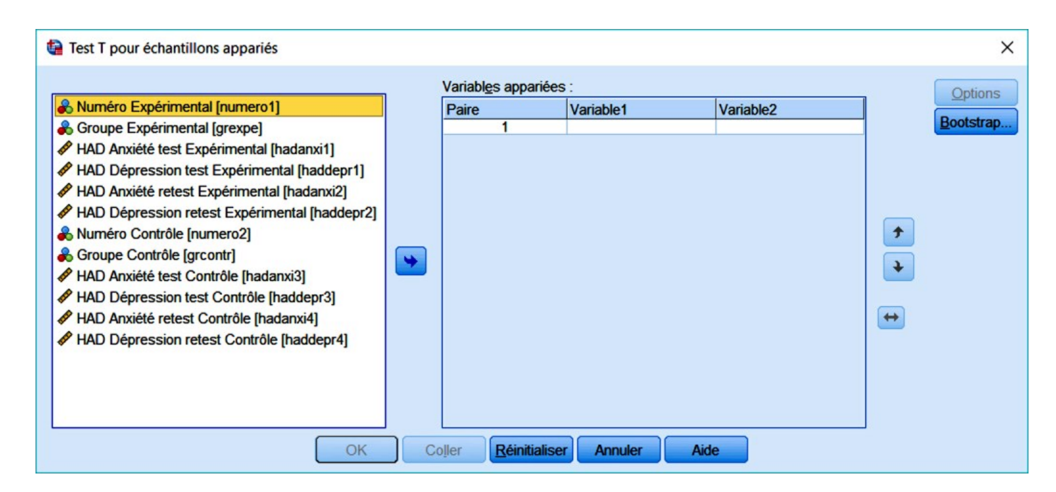


Figure 67 : Fenêtre du test T pour échantillons appariés.

Le jeu consiste à glisser dans la cellule « **Variables appariées :** » des paires de variables, c'est-à-dire la même

variable au test et au retest, cela respectivement pour le groupe « **Expérimental** » et pour le groupe « **Contrôle** ». Par exemple, vous cliquez pour sélectionner « **HAD Anxiété test Expérimental** » et, tout en tenant la touche **Ctrl** du clavier pressée, vous sélectionnez « **HAD Anxiété retest Expérimental** ». Puis, vous cliquez sur le bouton flèche bleu pour le transfert dans la cellule « **Variables appariées :** ».

Comme nous avons 2 VD (anxiété et dépression) pour 2 groupes (« **Expérimental** » et « **Contrôle** »), cela fait bien 4 paires de variables à programmer, soit :

- HAD Anxiété test Expérimental + HAD Anxiété retest Expérimental
- HAD Dépression test Expérimental + HAD Dépression retest Expérimental
- HAD Anxiété test Contrôle + HAD Anxiété retest Contrôle
- HAD Dépression test Contrôle + HAD Dépression retest Contrôle

Ce qui donne (Fig. 68) :

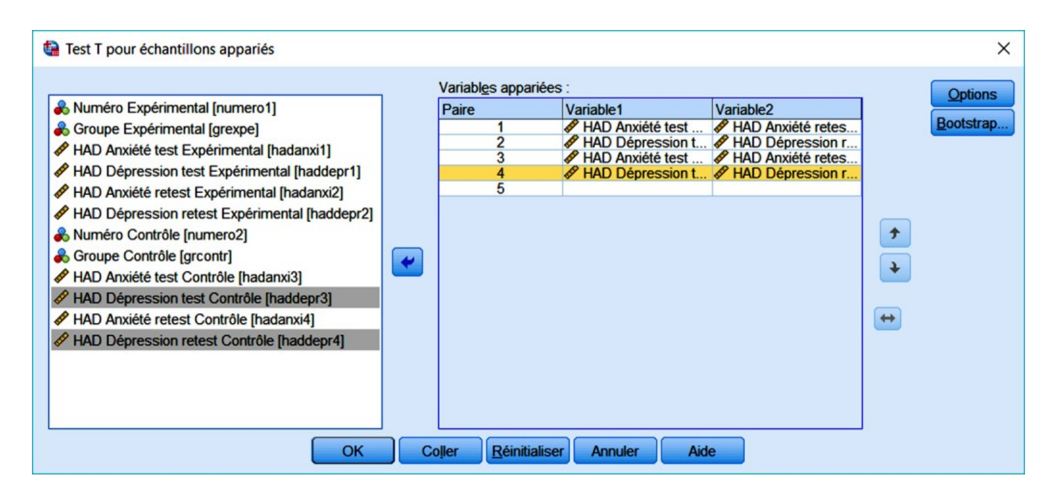


Figure 68 : Paramétrage de la fenêtre du test T pour échantillons appariés.

Il n'y a aucune option à paramétrer. Bouton **[OK]**, le résultat obtenu est le suivant (Tab. 21). Le test T s'interprète comme une ANOVA : le **t** doit être élevé et le ***P* =< 0,05** pour avoir une différence significative entre test et retest pour un même groupe. On observe une différence significative entre test et retest pour le groupe « **Expérimental** » à l'anxiété **(t(19) = 11,301 ; *P* = 0,000)**, comme à la dépression **(t(19) = 10,804 ; *P* = 0,000)**. Alors que pour le groupe de « **Contrôle** », il n'y a pas de différence : anxiété **(t(19) = 0,694 ; *P* = 0,496)**, dépression **(t(19) = -1,329 ; *P* = 0,200)**.

Remarque : (19) entre parenthèses indique (comme au F de l'ANOVA) les degrés de liberté (ddl).

Test des échantillons appariés

		Différences appariées							
					Intervalle de confiance de la différence à 95 %				
		Moyenne	Ecart type	Moyenne erreur standard	Inférieur	Supérieur	t	ddl	Sig. (bilatéral)
Paire 1	HAD Anxiété test Expérimental - HAD Anxiété retest Expérimental	11,550	4,571	1,022	9,411	13,689	11,301	19	,000
Paire 2	HAD Dépression test Expérimental - HAD Dépression retest Expérimental	9,150	3,787	,847	7,377	10,923	10,804	19	,000
Paire 3	HAD Anxiété test Contrôle - HAD Anxiété retest Contrôle	,500	3,220	,720	-1,007	2,007	,694	19	,496
Paire 4	HAD Dépression test Contrôle - HAD Dépression retest Contrôle	-1,150	3,870	,865	-2,961	,661	-1,329	19	,200

Tableau 21 : Résultats au test T pour échantillons appariés.

Procédure pour le test Z de Wilcoxon

Allez au menu **Analyse/Tests non paramétriques/Boîte de dialogue ancienne version/2 échantillons liés**. La fenêtre qui s'ouvre ressemble beaucoup à celle du test T et elle fonctionne de manière identique. Sur le même principe que le test T, vous placez les paires test-retest des variables dans la cellule « **Paires à tester :** », jusqu'à obtenir ceci (Fig. 69) :

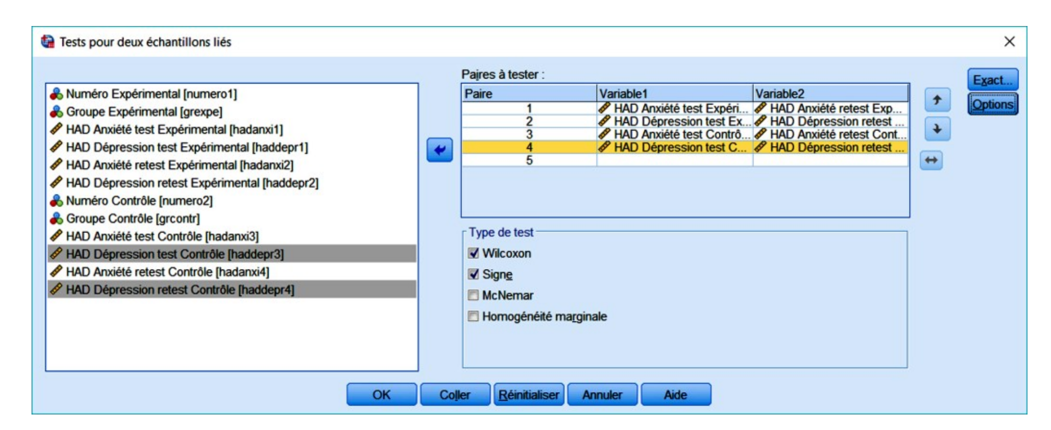


Figure 69 : Paramétrage de la fenêtre du test Z de Wilcoxon.

Tests statistiques[a]

	HAD Anxiété retest Expérimental - HAD Anxiété test Expérimental	HAD Dépression retest Expérimental - HAD Dépression test Expérimental	HAD Anxiété retest Contrôle - HAD Anxiété test Contrôle	HAD Dépression retest Contrôle - HAD Dépression test Contrôle
Z	-3,923[b]	-3,829[b]	-,570[b]	-1,225[c]
Sig. asymptotique (bilatérale)	,000	,000	,569	,221

a. Test de classement de Wilcoxon

b. Basée sur les rangs positifs.

c. Basée sur les rangs négatifs.

Tableau 22 : Résultats au test Z de Wilcoxon

Bouton **[Options]** vous cochez « **Caractéristique** », bouton

[Poursuivre]. Pour le « **Type de test** », cochez le « **Signe** » (test de signifiance) en plus du « **Wilcoxon** ». Bouton **[OK]**, cela donne (Tab. 22).

Le Z du test de Wilcoxon s'interprète, lui aussi, comme l'ANOVA : plus il est élevé et plus son coefficient de signification est bas et plus la différence entre les moyennes est significative. Dans notre cas de figure, nous avons pour le groupe « **Expérimental** » une différence significative entre test et retest à l'anxiété **(Z = -3,923 ; *P* = 0,000)** et à la dépression **(Z = -3,829 ; *P* = 0,000)**. Par contre, pas de différence significative pour le groupe de « **Contrôle** » : anxiété **(Z = -0,570 ; *P* = 0,569)**, dépression **(Z = -1,225 ; *P* = 0,221)**.

Pour en terminer avec les statistiques inférentielles et en ce qui concerne ce que vous devez présenter dans votre mémoire, quelques remarques. Dans le corpus du mémoire, vous ne mettez que le strict nécessaire pour présenter votre travail. Vous pouvez même ignorer les moyennes et ne présenter que les résultats aux tests statistiques (**χ2**, **ANOVA**, **corrélation de Pearson**, **test T** ou **Wilcoxon**). Vous réservez donc les tableaux détaillés fournis par SPSS pour les annexes auxquelles le lecteur peut se reporter pour en savoir plus, s'il le souhaite. Pour un article, faite comme pour le corpus du mémoire, tout le secret d'un bon article réside dans sa concision.

ANNEXES

SYNOPTIQUE DES MÉTHODES STATISTIQUES

Vous allez retrouver dans cette annexe, l'ensemble des 5 méthodes statistiques de base couramment utilisées pour des petites statistiques dans un mémoire ou un article « classique » en sciences humaines.

Pour sélectionner la bonne méthode demandez-vous quels types de variables vous allez tester. Vous avez 3 cas de figure :

1) Quantitative / Qualitative
2) Qualitative / Qualitative
3) Quantitative / Quantitative

Ensuite, dans le troisième cas de figure (Quantitative/ Quantitative), demandez-vous si vous êtes face à de petits échantillons (n < 30) ou non. À partir de cette analyse, il ne vous reste plus qu'à vous reporter aux intitulés des tableaux pour sélectionner la bonne méthode.

Par exemple : je veux tester une variable quantitative avec une variable qualitative, donc je choisis l'ANOVA. Et ainsi de suite…

Types de variables	**Quantitative**
Qualitative	**ANOVA** **Procédure :** Analyse/Comparer les moyennes/ANOVA à 1 facteur **Liste Variables dépendantes :** variables quantitatives **Facteur :** variable de regroupement **[Options]** : Caractéristiques **Résultat :** (F(1,38) = 0,506 ; *P* = 0,680) (« **(1,38)** » = degrés de liberté : Inter-groupes, Intra-groupes) **Interprétation :** différence significative si F élevé et *P* <= 0,05

Types de variables	**Qualitative**
Qualitative	**Khi-deux** **Procédure :** Analyse/Statistiques descriptives/Tableaux croisés **Ligne(s) :** variables **Colonne(s) :** variable de regroupement **[Statistiques]** : Khi-carré **Résultat :** (χ^2 = 8,565 ; *P* = 0,036) **Interprétation :** différence significative si χ^2 élevé et *P* <= 0,05

Types de variables	**Quantitative**
Quantitative	**Corrélation** **Procédure :** Analyse/Corrélation/Bivariée **Variables :** variables quantitatives **[Options]** : Moyennes et écarts-types **Résultat :** (r = 0,448 ; P = 0,004) **Interprétation :** corrélation si r >= 0,30 et P <= 0,05

Types de variables	**Quantitative**
Quantitative	**Test T de Student** (petits effectifs <=30 et distribution non normale) **Procédure :** Analyse/Comparer les moyennes/Test-T pour échantillons appariés **Variables appariées :** paires de variables (ex. : test-retest) **Résultat :** (t(19) = 10,804 ; P = 0,000) (« **(19)** » = degrés de liberté) **Interprétation :** différence significative si T élevé et P <= 0,05

Types de variables	**Quantitative**
Quantitative	**Test Z de Wilcoxon** (petits effectifs <=30 et distribution non normale) **Procédure :** Analyse/Tests non paramétriques/Boîte de dialogue ancienne version/2 échantillons liés **Variables appariées :** paires de variables (ex. : test-retest) **Type de test :** Wilcoxon et Signe **[Options]** : Caractéristique **Résultat :** (Z = -3,923 ; *P* = 0,000) **Interprétation :** différence significative si Z élevé et *P* <= 0,05

COMMENT OBTENIR SPSS ?

Entrez le terme « spss » dans Google. Cliquez sur le premier lien « ibm.com » qui s'affiche. Normalement, le lien est adapté à la langue de votre navigateur, donc le français, par exemple :[26]

https://www.ibm.com/fr-fr/spss

Cliquez sur le lien « Essayer gratuitement SPSS Statistics », ce qui vous conduit vers une page formulaire pour ouvrir un compte gratuit. Formulaire que vous devez remplir, en prenant garde à ce que votre adresse email soit authentique et fonctionnelle. Car c'est sur cette adresse email que vous allez recevoir le lien pour télécharger SPSS.

ATTENTION : ne faites pas cette procédure avant d'avoir préparé votre dossier Excel, car la durée de 30 jours de l'essai gratuit démarre à la date d'enregistrement de votre compte SPSS.

Et bonne chance pour votre mémoire ou article !

[26] Notez bien que les informations présentées dans cette annexe sont valables pour la date d'édition de l'ouvrage (juin 2024). Elles pourront varier au fil des années en fonction des stratégies de l'éditeur de SPSS Statistics.

À PROPOS DE L'AUTEUR

Le docteur Éric Loonis est psychologue clinicien et psychopathologue. Il a exercé en tant que psychothérapeute libéral auprès de familles, de couples et de personnes addictées. Il a mené une longue carrière de psychologue clinicien en institut médico-éducatif auprès d'enfants, d'adolescents et de jeunes adultes handicapés. Le docteur Éric Loonis est l'auteur de nombreux articles et ouvrages scientifiques sur les thèmes de l'addiction et de l'imaginaire érotique.

Un manuel de statistiques simple et sans aucune formule.

Vous vous sentez sous-doué en statistiques et vous pensez ne jamais y arriver ? Vous vous dites que pour un « littéraire » comme vous, un simple étudiant en sciences humaines, les stats c'est trop. Et pourtant, dès le master et plus tard en thèse, il vous faudra bien en faire, si vous voulez valider vos recherches et soutenir vos mémoires. Si vous êtes thésard ou docteur, il vous faut encore des statistiques pour vos publications, nul n'y échappe ! Alors cet ouvrage est là pour vous aider.

Ne pas être doué en stats n'est pas une fatalité. Les statistiques c'est un peu comme la cuisine : il n'est pas nécessaire de connaître la chimie pour pouvoir appliquer les bonnes recettes.

Un manuel de statistiques atypique, car il ne comporte ni formules, ni explications mathématiques absconses. Seulement des trucs, des recettes, des façons d'interpréter les résultats. Autant dire que c'est un ouvrage essentiellement pratique, comme un mode d'emploi. Il s'adresse aux étudiants de l'université (en particulier en sciences humaines). Cette nouvelle édition, comportant plus d'une centaine de tableaux et de copies d'écran, est entièrement révisée et mise à jour.

